THE ADVANCED

THEORY OF STATISTICS

THE ADVANCED THEORY OF STATISTICS

by

MAURICE G. KENDALL, M.A.

An Honorary Secretary of the Royal Statistical Society
Statistician to the Chamber of Shipping of the United Kingdom
Fellow of the Institute of Mathematical Statistics

VOLUME II

With 30 Illustrations and 52 Tables

LONDON

CHARLES GRIFFIN & COMPANY LIMITED

42 DRURY LANE

1946

TO

PETER AND PAUL

Printed in Great Britain
by Butler & Tanner Limited, Frome

PREFACE TO VOLUME II

This volume falls into five sections. The first, comprising chapters 17 to 20, deals with Estimation. The second, comprising chapters 21, 23, 24 and 26 to 28, covers the Theory of Statistical Tests, including the Analysis of Variance and Multivariate Analysis. The third, consisting of chapter 22, deals with Regression Analysis and completes the account of statistical relationship begun in chapters 13 to 16 of Volume I. In the fourth, chapter 25, I have tried to give an introductory account of the reaction of theoretical considerations on the Design of Statistical Inquiries. Finally, the fifth, comprising chapters 29 and 30, deals with the Analysis of Time-Series.

The literature of statistical theory is now so vast that it seemed worth while devoting considerable space to a bibliography, which is given in Appendix B. Although it is far from complete, I hope that it will serve its purpose in guiding the student to the main sources.

The chief problem in the writing of this volume arose in connection with the logic of statistical inference. Whenever possible I have kept the treatment objective. It is, I consider, unfair in a book of this kind not to present all sides of a case, particularly when there is so much disagreement among the authorities. Some day I hope to show that this disagreement is more apparent than real, and that all the existing theories of inference in probability differ essentially only in matters of taste in the choice of postulates. But this book is not the place for such work, and for the present I am content to state the position and to leave the reader to exercise his own choice.

The difficulty became most acute in dealing with confidence intervals and fiducial inference, where two approaches which at first sight appear identical can lead to different results. Rather than try to reconcile them I have written a separate chapter on each. Professor E. S. Pearson was kind enough to read the manuscript of chapter 19 and Professor R. A. Fisher that of chapter 20, so that I think their respective views are, at any rate, not misrepresented. I am very grateful to them both for their help in this connection.

My thanks are also due to Mr. P. A. Moran and Mr. A. J. H. Morrell, who cheerfully undertook to help with the proof reading and to whose painstaking scrutiny I owe the removal of a number of obscurities and errors. I shall be grateful to any reader who detects and notifies me of any further slips which have evaded us. Once again I have also to thank the publishers and the printers for the trouble they have taken in the production of the finished work.

M. G. K.

LONDON,
April, 1946.

TABLE OF CONTENTS

CHAPTER 17

ESTIMATION: LIKELIHOOD

The Problem

17.1. On several occasions in previous chapters we have encountered the problem of estimating from a sample the values of the parameters of the parent population. We have hitherto dealt on somewhat intuitive lines with such questions as arose—for example, in the theory of large samples we have taken the means and moments of the sample to be satisfactory estimates of the corresponding means and moments in the parent.

We now proceed to study this branch of the subject in more detail. In the earlier part of the present chapter we shall examine the sort of criteria which are required of a "good" estimate and discuss the question whether there exist "best" estimates in any acceptable sense of the term. In the remainder of the chapter and in Chapter 18 we shall consider various methods of obtaining estimates with the required properties. In Chapters 19 and 20 we shall look at the same problem from a rather different point of view and discuss the theories of confidence intervals and fiducial limits.

17.2. It will be evident that if a sample is not random and nothing precise is known about the nature of the bias operating when it was chosen, very little can be inferred from it about the parent population. Certain conclusions of a trivial kind are sometimes possible—for instance, if we take ten turnips from a pile of 100 and find that they weigh ten pounds altogether, the mean weight of turnips in the pile must be greater than one-tenth of a pound; but such information is rarely of value, and estimation based on biassed samples remains very much a matter of individual opinion and cannot be reduced to exact and objective terms. We shall therefore confine our attention to random samples only. Our general problem, in its simplest terms, is then to estimate the value of a parameter in the parent from the information given by the sample. In the first instance we consider the case when only one parameter is to be estimated. The case of several parameters will be discussed later.

17.3. Let us in the first place consider what we mean by "estimation". We know, or assume as a working hypothesis, that the parent population is distributed in a form which would be completely determinate if we knew the value of some parameter θ. We are given a sample of values $x_1 \ldots x_n$. We require to determine, with the aid of the x's, a number which can be taken to be the value of θ, or a range of numbers which can be taken to include that value.

Now a single sample, considered by itself, may be rather improbable, and any estimate based on it may therefore differ considerably from the true value of θ. It appears, therefore, that we cannot expect to find any method of estimation which can be guaranteed to give us a close estimate of θ on every occasion and for every sample. We must content ourselves with formulating a rule which will give good results "in the long run" or "on the average", or which has "a high probability of success"—phrases which express the fundamental fact that we have to regard our method of estimation as generating a population of estimates and to assess its merits according to the properties of this population.

17.4. It will clarify our ideas considerably if we draw a distinction between the method or rule of estimation, which, following Pitman, we shall call an Estimator, and the value to which it gives rise in particular cases, the Estimate. The distinction is the same as that between a function $f(x)$, regarded as defined for a range of the variable x, and the particular value which the function assumes, say $f(a)$, for a specified value of x equal to a. Our problem is not to find estimates, but to find Estimators. We do not reject a method because it gives a bad result in a particular case (in the sense that the estimate differs materially from the true value). We should only reject it if it gave bad results in the long run, that is to say, if the population of possible values of the estimator were seriously discrepant with the value of θ. The merit of the estimator is judged by the population of estimates to which it gives rise. It is itself a random variable and has a distribution to which we shall frequently have occasion to refer.

17.5. In the theory of large samples we have often taken as an estimator of a parameter θ a statistic t calculated from the sample in exactly the same way as θ is calculated from the population, e.g. the sample-mean is taken as an estimate of the parent mean. Let us examine how this procedure can be justified. Consider the case when the parent population is

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}(x-\theta)^2\right\} dx, \qquad -\infty \leqslant x \leqslant \infty . \qquad (17.1)$$

Requiring an estimator for the parent mean θ, we take

$$t = \frac{1}{n}\sum_{j=1}^{n} x_j. \qquad (17.2)$$

The distribution of t is

$$dF = \frac{\sqrt{n}}{\sqrt{(2\pi)}} \exp\left\{-\frac{n}{2}(t-\theta)^2\right\} dt, \qquad (17.3)$$

that is to say, t is distributed normally about θ with variance $1/n$. We notice two things about this distribution : (a) it has a mean (and median and mode) at the true value θ, and (b) as n increases, the scatter of possible values of t about θ becomes smaller, so that the probability that a given t differs by more than a fixed amount from θ decreases. We may say that the accuracy of the estimator increases as n increases, or simply with n.

17.6. Generally, it will be clear that the phrase "accuracy increasing with n" has a definite meaning whenever the sampling distribution of t has a variance which decreases with $1/n$ and a central value which is either identical with θ or differs from it by a quantity which also decreases with $1/n$. Many of the estimators with which we are commonly concerned are of this type, but there are exceptions. Consider, for example, the Cauchy population

$$dF = \frac{1}{\pi}\,\frac{dx}{1+(x-\theta)^2}, \qquad -\infty \leqslant x \leqslant \infty . \qquad (17.4)$$

The mean (assuming that we conventionally agree that it exists) is at $x = \theta$. But if we try to estimate θ by the mean-statistic t we have, for the distribution of t,

$$dF = \frac{1}{\pi}\,\frac{dt}{1+(t-\theta)^2}, \qquad -\infty \leqslant t \leqslant \infty . \qquad (17.5)$$

(Cf. Example 10.1, vol. I, pp. 233–4.) In this case the distribution of t is the same as that of any single value of the sample, and does not increase in accuracy as n increases.

Consistence

17.7. The property of possessing increasing accuracy is evidently a very desirable one; and indeed, if the variance of the sampling distribution decreases with increasing n it is *necessary* that its central value should tend to θ, for otherwise the estimator would have values differing systematically from the true value and would be useless, not to say dangerous. We therefore formulate our first criterion for a suitable estimator as follows :—

An estimator t_n, computed from a sample of n values, will be said to be a consistent estimator of θ if, for any positive ε and η, however small, there is some N such that the probability that

$$|t_n - \theta| < \varepsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.6)$$

is greater than $1 - \eta$ for all $n > N$. In the notation of the theory of probability,

$$P\{|t_n - \theta| < \varepsilon\} > 1 - \eta, \qquad n > N. \quad . \quad . \quad . \quad (17.7)$$

The definition bears an obvious analogy to the definition of convergence in the mathematical sense. Given any fixed small quantity ε we can find a large enough sample number such that for all samples over that size the probability that t differs from the true value by more than ε is as near zero as we please. t_n is said to *converge in probability* to θ. Thus t is a consistent estimate of θ if it converges to θ in probability.

Example 17.1

The sample mean is a consistent estimator of the parameter θ in the population (17.1). This we have already established in general argument, but more formally the proof would proceed as follows :—

Suppose we are given ε. From (17.3) we see that $(t - \theta)\sqrt{n}$ is distributed normally about zero with unit variance. Thus the probability that $|(t - \theta)\sqrt{n}| \leqslant \varepsilon\sqrt{n}$ is the value of the normal integral between limits $\pm\varepsilon\sqrt{n}$. Given any positive η, we can always take n large enough for this quantity to be greater than $1 - \eta$ and it will continue to be so for any larger n. N may therefore be determined and the inequality (17.7) is satisfied.

Example 17.2

Suppose we have a statistic t_n whose mean value differs from θ by order n^{-1}, whose variance v_n is of order n^{-1} and which tends to normality as n increases. Clearly $(t_n - \theta)/\sqrt{v_n}$ will then tend to zero in probability and t_n will be consistent. This covers a great many statistics encountered in practice.

Unbiassed Estimators

17.8. The property of consistence is a limiting property, that is to say, it concerns the behaviour of an estimator as the sample number tends to infinity. It requires nothing of the behaviour for finite n, and if there exists one consistent estimator t_n we may construct infinitely many others; e.g.

$$\frac{n-a}{n-b}t_n$$

is also consistent. We have seen that in some circumstances a consistent estimator of the mean is the sample mean

$$\bar{x} = \frac{1}{n}\Sigma x_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.8)$$

But so is

$$\bar{x}' = \frac{1}{n-1} \Sigma x_j. \qquad (17.9)$$

Why do we prefer one to the other ? Intuitively it seems absurd to divide the sum of n quantities by anything other than their number n. We shall see in a moment, however, that intuition is not a very reliable guide on such matters. There are reasons for preferring

$$\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2 \qquad (17.10)$$

to

$$\frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 \qquad (17.11)$$

as an estimator of the parent variance, notwithstanding that the latter is the sample variance.

17.9. Consider the sampling distribution of an estimator t. If the estimator is consistent, its distribution must, for large samples, have a central value in the neighbourhood of θ. We may choose among the field of consistent estimators by requiring that θ shall be equated to this central value not merely for large, but for all samples. Whether we choose as the appropriate central value the mean, the median or the mode is to some extent a matter of taste. We shall consider below what follows if we select the mode (which gives us the maximum likelihood estimators). For the present we discuss the mean.

If we require that for all n the mean value of t shall be θ, we define what is known as an *unbiassed* estimator :

$$E(t) = \theta. \qquad (17.12)$$

This is an unfortunate word, like so many in statistics. There is nothing except convenience to exalt the arithmetic mean above other measures of location as a criterion of bias. We might equally well have chosen the mode as determining the " unbiassed " estimator, in which case the mean estimator would be " biassed " whenever it gave a different result. Since the use of " unbiassed " in connection with the mean is fairly widespread, however, we shall continue to use it.*

Example 17.3

Since

$$E\left\{\frac{1}{n}\Sigma(x)\right\} = \frac{1}{n}\Sigma\{E(x)\}$$
$$= \frac{1}{n}\Sigma\mu_1 = \mu_1,$$

the mean-statistic is an unbiassed estimator of the parent mean whenever the latter exists. But the sample-variance is not an unbiassed estimator of the parent variance. We have

$$\begin{aligned} E\{\Sigma(x-\bar{x})^2\} &= E\left\{\Sigma\left[x - \frac{1}{n}\Sigma(x)\right]^2\right\} \\ &= E\left\{\frac{n-1}{n}\Sigma(x^2) - \frac{1}{n}\Sigma(x_j x_k)\right\}, \qquad j \neq k \\ &= (n-1)\mu_2' - (n-1)\mu_1'^2 \\ &= (n-1)\mu_2. \end{aligned}$$

* The word has already occurred in vol. I, p. 200, in this sense. It may be spelt with either one or two s's. My usage, I am afraid, is not consistent, but in this volume I use two.

Thus $\frac{1}{n} \Sigma (x - \bar{x})^2$ has a mean value $\frac{n-1}{n} \mu_2$. On the other hand, an unbiassed estimator is given by

$$\frac{1}{n-1} \Sigma (x - \bar{x})^2,$$

and for this reason it is sometimes preferred to the sample variance. There are other reasons which will appear when we come to study the analysis of variance.

Efficient Estimators

17.10. In general there will exist more than one consistent estimator of a parameter, even if we confine ourselves only to unbiassed estimators. Consider once again the estimation of the mean of a normal population with known variance. The sample mean is consistent and unbiassed. We will now prove that the same is true of the median.

Consideration of symmetry is enough to show that the median is an unbiassed estimate of the parent mean, which is, of course, the same as the parent median. For large n the distribution of the median tends to the normal form (cf. Example 9.7, vol. I, p. 213),

$$dF \propto \exp \{- 2nf_1^2 (x - \theta)^2\} dx \quad . \quad . \quad . \quad . \quad (17.13)$$

where f_1 is the median ordinate of the parent, in our present case $1/\sqrt{(2\pi)} = 0{\cdot}3989$. The variance tends to zero and the estimator is consistent. Its variance is $\pi/2n$.

17.11. We are therefore at liberty to seek for further criteria to choose between estimators with the common property of consistence. Such a criterion arises naturally if we consider the sampling variances of the estimators. Generally speaking, the estimator with the smaller variance will be grouped more closely round the value θ ; this will certainly be so for distributions of the normal type. An estimator with a smaller variance will therefore deviate less, on the average, from the true value than one with a larger variance. Hence we may reasonably regard it as better or more *efficient.*

If, of two consistent estimators t_1 and t_2, we have var $t_1 <$ var t_2 for all n, then t_1 is more efficient than t_2 for all sample sizes. It is possible to have var $t_1 <$ var t_2 for some ranges of n and var $t_1 >$ var t_2 for others, in which case the estimators are more or less efficient in different ranges.

In the case of mean and median we have, for any n,

$$\text{var (mean)} = \frac{\sigma^2}{n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.14)$$

and for large n

$$\text{var (median)} = \frac{\pi\sigma^2}{2n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.15)$$

where σ^2 is the parent variance. Since $\pi/2 = 1{\cdot}57 > 1$ the mean is more efficient than the median for large n at least. For small n we have to work out the variance of the median. The following values may be obtained from those given in Table XXIII of *Tables for Statisticians and Biometricians, Part II* :—

n	2	3	4	5
var (median)	1·00	1·35	1·19	1·44

It appears that the mean is always more efficient than the median in estimating the parameter θ for the normal distribution (17.1).

Example 17.4

For the Cauchy distribution

$$dF = \frac{1}{\pi} \frac{dx}{1 + (x - \theta)^2}, \qquad -\infty \leqslant x \leqslant \infty$$

we have already seen that the sample mean is not a consistent estimator. However, for the median in large samples we have, since the median ordinate is $1/\pi$,

$$\text{var (median)} = \frac{\pi^2}{4n}.$$

It is seen that the median is consistent, and although direct comparison with the mean is not possible because the latter does not possess a sampling variance, the median is evidently a better estimator for θ than the mean. This provides an interesting contrast with the case of the normal parent, particularly in view of the similarity of the parent frequency-distributions.

17.12. In some cases, as we shall see below, there exist consistent estimators whose sampling variance for large samples is less than that of any other such estimator. We shall call such estimators most-efficient. When they exist they provide a standard of measurement of efficiency. In fact, if t_2 has variance v_2 and the most-efficient estimator t_1 has variance v_1, the efficiency E of t_2 is defined as

$$E = \frac{v_1}{v_2}. \qquad (17.16)$$

It will be seen later that in normal samples the mean is a most-efficient estimator, so that the efficiency of the median for such samples is

$$= \frac{2n}{\pi} \cdot \frac{1}{n} = 0{\cdot}637.$$

17.13. If we have a sample of 100 members the variance of the median (assuming normality) will be about the same as that of the mean in only 64 members. Thus, if sampling variance be accepted as a criterion of accuracy of estimation, the use of the median instead of the mean sacrifices about 36 observations in 100. It is not possible to economise by using a different estimator than the mean.

Other things being equal, the estimator with the greater efficiency is undoubtedly the one to use. But sometimes other things are not equal. It may, and does, happen that a most-efficient estimate derived from t_1 is more troublesome to calculate than an alternative t_2. The extra labour involved in calculation may be greater than the saving in dealing with a smaller sample number, particularly if there are plenty of further observations to hand.

Example 17.5

Consider the estimation of the standard deviation of a normal population with variance σ^2 and unknown mean. Two possible estimators are the standard deviation of the sample (or the square-root of $\Sigma (x - \bar{x})^2/(n - 1)$ if it is desired to use an unbiassed estimator) and the mean deviation of the sample multiplied by $\sqrt{(\pi/2)}$ (cf. **5.20**). The latter is easier to calculate, as a rule, and if we have plenty of observations (as, for example, if we are finding the standard deviation of a set of barometric records and the addition of further

members to the sample is merely a matter of turning up more records) it may be worth while estimating from the mean-deviation rather than from the standard deviation.

In normal samples the variance of the mean-deviation is (**9.13**)—

$$\frac{2}{\pi}\frac{n-1}{n^2}\sigma^2\left(\frac{\pi}{2}+\sqrt{\{n(n-2)\}}-n+\arcsin\frac{1}{n-1}\right)\sim\frac{\sigma^2}{n}\left(1-\frac{2}{\pi}\right). \quad (17.17)$$

The variance of the estimator from the mean deviation is then approximately

$$\frac{\sigma^2}{n}\left(\frac{\pi-2}{2}\right). \quad (17.18)$$

Now the variance of the standard deviation is (**9.22**) $\sigma^2/2n$, and we shall see later that it is a most-efficient estimator. Thus the efficiency of the first estimator is

$$E=\frac{\sigma^2}{2n}\bigg/\frac{\sigma^2}{n}\left(\frac{\pi-2}{2}\right)=\frac{1}{\pi-2}=0{\cdot}876.$$

The accuracy of the estimate from the mean deviation of a sample of 1000 is then about the same as that from the standard deviation of a sample of 876. If it is easier to calculate the m.d. of 1000 observations than the s.d. of 876 and there is no shortage of observations, it may be more convenient to use the former.

It has to be remembered, nevertheless, that in adopting such a procedure we are deliberately wasting information. By taking greater pains we could improve the efficiency of our estimate from 0·876 to unity, or by about 14 per cent. of the former value.

Sufficient Estimators

17.14. The comparison of the efficiencies of two estimators, as measured by their variances, may be made for any n, but the absolute efficiency as defined in **17.12** by relation to a most-efficient estimator is in the main a limiting property. We shall see below (**17.36**) that the definition may be extended to small samples and to non-normal variation, but most-efficient estimators for finite n do not exist so frequently in statistical practice as in the limiting case of large samples. Sometimes, however, there are estimators which may be regarded as the " best " for samples of any size, and we proceed to consider them.

Before doing so, we prove that, in the limit, all most-efficient estimators tend to equivalence.

More precisely, if two most-efficient estimators t_1 and t_2 tend in the limit to be distributed in the bivariate form

$$dF\propto\exp\left[-\frac{1}{2v(1-\rho^2)}\{(t_1-\theta)^2-2\rho\,(t_1-\theta)\,(t_2-\theta)+(t_2-\theta)^2\}\right]dt_1\,dt_2, \quad (17.19)$$

then the correlation $\rho=1$. Here v is the variance of each estimator.

Consider the estimator

$$u_1=\tfrac{1}{2}\,(t_1+t_2).$$

Clearly u_1 is consistent since t_1 and t_2 are both so. Putting

$$u_2=\tfrac{1}{2}\,(t_1-t_2)$$

we have, for the joint distribution of u_1 and u_2,

$$dF\propto\exp\left[-\frac{1}{2v\,(1-\rho^2)}\{2\,(1-\rho)\,(u_1-\theta)^2+2\,(1+\rho)\,u_2^2\}\right]du_1\,du_2. \quad (17.20)$$

Thus u_2 is distributed independently of u_1 and θ and we have

$$\text{var } u_1 = \frac{v\,(1-\rho^2)}{2\,(1-\rho)} = \frac{1+\rho}{2}\,v. \qquad (17.21)$$

Now t_1 is a most-efficient estimator and hence

$$\frac{1+\rho}{2}\,v = \text{var } u_1 \geqslant \text{var } t_1 = v$$

giving

$$\frac{1+\rho}{2} \geqslant 1. \qquad (17.22)$$

But ρ cannot be greater than unity and hence $\rho = 1$, which proves the theorem.

17.15. Consider once again the estimation of θ in the normal population (17.1). The joint distribution of the sample is given by

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{n}(x_j-\theta)^2\right\} dx_1 \ldots dx_n \qquad (17.23)$$

We have the familiar result

$$\sum_{j=1}^{n}(x_j-\theta)^2 = \Sigma\,(x-\bar{x})^2 + n\,(\bar{x}-\theta)^2,$$

and hence

$$dF = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{n}{2}(\bar{x}-\theta)^2\right\} \exp\left\{-\tfrac{1}{2}\,\Sigma\,(x-\bar{x})^2\right\} dx_1 \ldots dx_n \qquad (17.24)$$

Thus the frequency function of the distribution of x's (which is equivalent to the likelihood function) can be factorised into two parts, one depending on $\bar{x}$ and θ, the other depending on the x's but not on θ.

The quantity $\bar{x}$ is then said to be a *sufficient* estimator of θ; and generally, if the likelihood function is expressible in the form (as a product of two frequency functions)—

$$L\,(x_1, \ldots x_n, \theta) = L_1\,(t, \theta)\, L_2\,(x_1, \ldots x_n), \qquad (17.25)$$

where L_1 does not contain the x's otherwise than in the form t and L_2 is independent of θ, t is said to be a sufficient estimator of θ.

17.16. As so defined, a sufficient estimator, if it exists at all, is unique except that if t obeys the relation (17.25) any function of t will obviously also obey the same relation. From all such functions we must evidently choose one which gives a consistent estimator and can sometimes, as in the example of the previous section, find the estimator which is unbiassed. Apart from such ambiguities, which offer no difficulties in practice, the property of uniqueness holds. For if t_1 and t_2 were two different sufficient statistics, not functionally related, we should have—

$$L_1\,(t_1, \theta)\, L_2\,(x_1, \ldots x_n) \equiv M_1\,(t_2, \theta)\, M_2\,(x_1, \ldots x_n),$$

and hence

$$\frac{L_1\,(t_1, \theta)}{M_1\,(t_2, \theta)} \equiv \frac{M_2}{L_2}. \qquad (17.26)$$

Since the expression on the right does not contain θ, L_1 must be a factor of M_1 and moreover the quotient must be a constant; for if it were a function of the x's that function would have been assimilated to L_2 or M_2.

Hence

$$L_1(t_1, \theta) \equiv k\, M_1(t_2, \theta),$$

and this cannot be so unless t_1 and t_2 are functionally related.

17.17. The fundamental property of sufficient estimators derives from the following theorem :—

If t_1 is sufficient and t_2 is any other estimator of θ (not a function of t_1) the joint distribution of t_1 and t_2 may be put in the form

$$dF = f_1(t_1, \theta)\, f_2(t_2, t_1)\, dt_1\, dt_2, \qquad (17.27)$$

where f_2 does not contain θ. Conversely, if (17.27) holds for every t_2 then t_1 is sufficient.

Before proving this result let us notice its importance. From (17.27) it follows that for any given t_1 the distribution of t_2 is equal to $f_2(t_2, t_1)\, dt_2$, i.e. is independent of θ. Consequently, if we know t_1, the probability of any range of values of t_2 is the same for all θ. The distribution of t_2 given t_1, therefore, can throw no light whatever on θ. Thus, a knowledge of t_1 gives all the information that the sample can supply about θ and no other estimator can add anything to it. We are clearly justified in such circumstances in describing a sufficient estimator as the " best ".

Now as to the theorem itself. The direct part is easily proved. In fact, we have from (17.25)—

$$L(x_1, \ldots x_n, \theta)\, dx_1 \ldots dx_n = L_1(t_1, \theta)\, L_2(x_1, \ldots x_n)\, dx_1 \ldots dx_n.$$

Make the transformation

$$\left.\begin{aligned} y_1 &= t_1(x_1, \ldots x_n) \\ y_2 &= t_2(x_1, \ldots x_n) \\ y_3 &= x_3 \\ &\cdot \quad \cdot \quad \cdot \\ y_n &= x_n \end{aligned}\right\} \qquad (17.28)$$

The element of frequency becomes

$$L_1(t_1, \theta)\, L_2(x_1, \ldots x_n)\, \frac{\partial(x_1, x_2)}{\partial(t_1, t_2)}\, dy_1 \ldots dy_n \qquad (17.29)$$

where the t's and x's are to be expressed in terms of the y's. We have excluded the case when t_2 is functionally related to t_1, and hence the Jacobian $\partial(x_1, x_2)/\partial(t_1, t_2)$ does not vanish identically. The frequency element of y_1 and y_2 is then obtained from (17.29) by integrating out the other variables. Since y_1 and y_2 are equal respectively to t_1 and t_2 this process will leave unchanged the function $L_1(t_1, \theta)$ and reduce the other part to a function of t_1 and t_2, say $f_2(t_1, t_2)$. Writing f_1 for L_1 we then have

$$dF = f_1(t_1, \theta)\, f_2(t_1, t_2)\, dt_1\, dt_2,$$

as stated in the theorem.

The converse is a little more difficult. Let t_1 be sufficient and make the transformation $y_1 = t_1$, $y_2 = x_2$, etc. The joint distribution of sample values becomes

$$L(x_1, \ldots x_n) = L'(t_1, y_2, \ldots y_n) \left| \frac{\partial t_1}{\partial x_1} \right| \qquad (17.30)$$

Since t_1 is independent of θ, so is $\partial t_1/\partial x_1$. Hence, if the distribution of t_1 is $f(t_1)\, dt_1$, L' may be written

$$f(t_1)\, L''(t_1, y_2, \ldots y_n), \qquad (17.31)$$

and the converse will be established if we can show that L'' does not contain θ. This we

do by demonstrating that if there are values $y_2' \ldots y_n'$ for which L'' assumes different values for different values of θ then the joint distribution of t_1 and t_2 cannot be independent of θ, which contradicts our hypothesis.

Suppose, then, that for two values of θ, say θ_1 and θ_2,

$$L''(t_1, y_2', \ldots y_n')_{\theta_1} = L''(t_1, y_2', \ldots y_n')_{\theta_2} + 2\alpha, \qquad (17.32)$$

where α is not zero. Consider a new statistic t_3 defined by

$$t_3^2 = \sum_{j=2}^{n} (y_j - y_j')^2 \qquad (17.33)$$

Assuming that L'' is continuous in the y's, we may determine a value of t_3, say t_3', such that

$$L''(t_1, y_2, \ldots y_n)_{\theta_1} \geqslant L''(t_1, y_2, \ldots y_n)_{\theta_2} + \alpha \qquad (17.34)$$

everywhere inside the range of values bounded by

$$t_3'^2 = \Sigma (y - y')^2.$$

Then for any fixed t_1 the total frequency inside this range is obtained by integrating L'' over the appropriate values, and we shall find, in virtue of (17.34),

$$f_{\theta_1} > f_{\theta_2}, \qquad (17.35)$$

the f's referring to total frequencies.

But if the joint distribution of t_1 and t_2 is

$$dF = h(t_1, t_2)_\theta \, dt_1 \, dt_2$$

we have for the frequencies f,

$$f_{\theta_1} = \int_0^{t_3'} h(t_1, t_2)_{\theta_1} \, dt_2$$

$$f_{\theta_2} = \int_0^{t_3'} h(t_1, t_2)_{\theta_2} \, dt_2$$

and hence

$$\int_0^{t_3'} \{h(t_1, t_2)_{\theta_1} - h(t_1, t_2)_{\theta_2}\} \, dt_2 > 0,$$

so that the joint distribution cannot be independent of θ.

The above demonstration relates to the case when the frequency functions are continuous. In the discontinuous case the argument simplifies and we leave it to the reader to supply the proof.

17.18. We now prove an important further result to the effect that a sufficient estimator is most-efficient, provided that a most-efficient estimator exists. We assume that the joint distribution of the sufficient estimator t_1 and any other estimator t_2 tends to normality for large n, say in the form

$$dF \propto \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(t_1-\theta)^2}{v_1} - \frac{2\rho(t_1-\theta)(t_2-\theta)}{\sqrt{(v_1 v_2)}} + \frac{(t_2-\theta)^2}{v_2}\right\}\right] dt_1 \, dt_2 \,. \qquad (17.36)$$

where v_1 and v_2 are the variances of t_1 and t_2 respectively. Since t_1 is sufficient, the distribution of t_2 given t_1 does not contain θ. Now the distribution of t_1 is

$$dF \propto \exp\left\{-\tfrac{1}{2}\frac{(t_1-\theta)^2}{v_1}\right\} dt_1 \qquad (17.37)$$

and hence that of t_2 given t_1 is

$$dF \propto \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(t_1-\theta)^2}{v_1} - \frac{2\rho(t_1-\theta)(t_2-\theta)}{\sqrt{(v_1 v_2)}} + \frac{(t_2-\theta)^2}{v_2}\right\} + \tfrac{1}{2}\frac{(t_1-\theta)^2}{v_1}\right] dt_2$$

which reduces to

$$dF \propto \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{\rho(t_1-\theta)}{\sqrt{v_1}} - \frac{(t_2-\theta)}{\sqrt{v_2}}\right\}^2\right] dt_2 \quad . \quad . \quad (17.38)$$

If this is not to involve θ we must have

$$\rho = \sqrt{\frac{v_1}{v_2}} = \sqrt{E}, \text{ where } E \text{ is the efficiency of } t_2. \quad . \quad (17.39)$$

Since $\rho \leqslant 1$ it follows that $v_1 \leqslant v_2$, i.e. t_1 has a smaller variance than any other estimator. Consequently, if there exists a most-efficient statistic, t_1 itself is most-efficient.

17.19. The criterion of sufficiency is not a limiting property. A sufficient estimator is best for any sample size since it gives all the information about θ that the sample can give; and it is most-efficient for large samples. If we could always find a sufficient estimator our problem would be solved, but unfortunately sufficiency is the exception rather than the rule.

Example 17.6

The frequency element of a sample of n from the population

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\} dx$$

can be put in the form

$$dF = \frac{\sqrt{n}}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{n}{2}\frac{(\bar{x}-m)^2}{\sigma^2}\right\} \frac{n^{\frac{n-1}{2}}}{(2\sigma^2)^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{ns^2}{2\sigma^2}} s^{n-3}\, d\bar{x}\, ds^2$$

(Cf. Example 10.5, vol. I, p. 238.)

If we know σ, then, as we have already seen, $\bar{x}$ is sufficient for m. But if we know m, s is *not* sufficient for σ. In fact, the factorisation in the above equation requires the appearance of σ in the element relating to $\bar{x}$, and we cannot separate a factor containing s and σ alone or the remaining variables alone.

This is what we might expect. If we know the real mean m there is little point in preferring the sample variance

$$s^2 = \frac{1}{n}\Sigma(x-\bar{x})^2$$

to the second moment

$$s'^2 = \frac{1}{n}\Sigma(x-m)^2$$

as an estimator of the parent variance. The distribution of s' is given by

$$dF = \frac{n^{\frac{n}{2}}}{(2\sigma^2)^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} e^{-\frac{ns'^2}{2\sigma^2}} (s')^{n-2}\, ds'^2$$

and this embodies the whole of the frequency element of the sample, apart from differentials in the other variables. Thus s' is sufficient for σ.

17.20. This completes the first stage of our inquiry. The criteria of consistence, efficiency and sufficiency provide standards which we shall look for in "good" estimators. Of themselves, however, they do not provide any systematic way of deriving estimators which obey them. We shall now consider various methods which have been proposed for providing estimators and examine how far they conform to our criteria. The most important method is that of maximum likelihood, which will occupy the remainder of this chapter. In the next chapter we shall consider four others, the method of minimum variance, the method of minimum χ^2, the method of least squares, and the method of inverse probability.

Maximum Likelihood

17.21. If the frequency function of the parent population is $f(x, \theta)$, the likelihood function of a sample of n is, by definition,

$$L = f(x_1, \theta)\, f(x_2, \theta) \ldots f(x_n, \theta) . \qquad (17.40)$$

The Principle of Maximum (or Maximal) Likelihood then states that if there exists a statistic $t = t(x_1, \ldots x_n)$ which maximises L for variations of θ, then t is to be taken as an estimator of θ. In short, t is the solution (if any) of

$$\frac{\partial L}{\partial \theta} = 0, \qquad \frac{\partial^2 L}{\partial \theta^2} < 0. \qquad (17.41)$$

Since L is positive, the first equation is equivalent to

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = \frac{\partial}{\partial \theta} \log L = 0, \qquad (17.42)$$

a form which is frequently more convenient.

There is one small point to notice here. In our usual convention, if a frequency function has a finite range, we regard it as defined from $-\infty$ to $+\infty$ but as zero outside that range. In this chapter we shall occasionally meet the reciprocal of f, which is undefined for zero f. Unless the contrary is specified we shall suppose that where f is zero $1/f$ is also to be regarded as zero. This will enable us to continue to regard the range as infinite, but some care is necessary where f is assumed everywhere continuous, for discontinuities may appear in f and $1/f$ at the terminals of the finite range. The point becomes important when we try to make certain existence theorems rigorous.

17.22. In sections **7.27** to **7.31** we touched on the principle of maximum likelihood from the point of view of statistical logic. We pointed out that its adoption required a new postulate in the theory of inference, but referred to the fact that the principle was recommended by the statistical properties of the estimators to which it leads. We now proceed to prove a series of theorems about these estimators, from which it will be seen that the posterior recommendation, so to speak, is very strong. In fact, maximum likelihood estimators are consistent, tend to normality for large n, have minimum variance in the limit at least, and provide sufficient statistics where such exist.

17.23. The reader may feel convinced intuitively that maximum likelihood estimators

are consistent, in which case he can pass to the next section. We shall now prove the result formally.

(a) If the frequency function $f(x, \theta)$ is continuous in x throughout its range, and

(b) if $f(x, \theta)$ is continuous and monotonic in θ in some θ-interval containing the true value of θ, say θ_0, and for all x in some x-interval,

then the maximum likelihood estimator of θ, say t, is consistent.

Our proof will also cover the case of discontinuous variates which can be reduced to the continuous case by replacing each value by an interval in which the frequency is uniformly distributed.

We first eliminate an inconvenience due to the infinitude of the range. In fact, if the range is infinite we make the variate transformation $x = \tan y$. The conditions (a) and (b) remain true of y, and the maximum likelihood estimator in x transforms to that in y. We may therefore take the range as finite.

The next step is to reduce the case to one of grouped frequencies by dividing the range into m intervals, the width of the jth interval being l_j. (We shall decide on the actual values of the l's below.) Writing

$$f_j = \int l_j f(x, \theta)\, dx, \qquad . \quad . \quad . \quad . \quad . \quad (17.43)$$

we have, in virtue of the continuity of f in x, that f_j/l_j differs as little as we please from $f(x_j, \theta)$. Then if L' is the likelihood of the grouped data, proportional to

$$\left(\frac{f_1}{l_1}\right)^{n_1} \left(\frac{f_2}{l_2}\right)^{n_2} \cdots \left(\frac{f_m}{l_m}\right)^{n_m} \qquad . \quad . \quad . \quad . \quad (17.44)$$

where n_j is the number of observations in the jth interval, we have, except for constants,

$$\log L' = \sum_{j=1}^{m} n_j \log f_j - \sum_{j=1}^{m} n_j \log l_j \qquad . \quad . \quad . \quad . \quad (17.45)$$

and this will differ arbitrarily little from the logarithm of the true likelihood

$$\log L = \sum_{j=1}^{n} \log f(x_j, \theta), \qquad . \quad . \quad . \quad . \quad . \quad (17.46)$$

provided that we take m large enough and the l's in consequence small enough.

Hence we see that if t is the estimator which maximises L and t' that which maximises L', in virtue of hypothesis (b) that L and L' are continuous in θ, t and t' will differ as little as we please for any given values of the x's and that uniformly. We may therefore prove our theorem for the finite number of variables n_j and infer its truth for the continuous case by proceeding to the limit.

In different samples the n_j will vary, subject only to the condition that $\Sigma(n_j) = n$. Let us choose the ranges l_j such that $f_j(\theta_0) = 1/m$ for all j, that is to say, such that the frequencies in all intervals are equal when θ takes its true value θ_0. Consider the likelihood function

$$K = \sum_{j=1}^{m} n_j \log z_j, \qquad . \quad . \quad . \quad . \quad . \quad (17.47)$$

where the z's are subject only to the condition

$$\Sigma(z) = 1. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (17.48)$$

We consider three values of K defined by particular values of the z's.

(a) When $z_j = n_j/n$, K is a maximum, say K_R. For we have

$$\delta K = \Sigma \frac{n_j}{z_j} \delta z_j$$

$$0 = \Sigma\, \delta z_j,$$

and hence

$$\frac{n_1}{z_1} = \frac{n_2}{z_2} = \ldots = \frac{\Sigma (n)}{\Sigma (z)} = n.$$

(b) When $z_j = f_j(\theta_0) = 1/m$, K is, say, K_M.

(c) When the estimator t' assumes the value, say, t_0' corresponding to the n_j's, and hence $z_j = f_j(t_0')$, K is a maximum, say K_Z, among the particular set of values of θ for which $z_j = f_j(\theta)$; for this is our definition of t'.

We have at once that

$$K_R \geqslant K_Z \geqslant K_M. \qquad \ldots \qquad (17.49)$$

Now, as the sample increases, the observed n_j/n converge in probability to their theoretical values $f_j(\theta_0) = 1/m$. Since K is continuous in the z's, $K_R - K_M$ will converge to zero in probability and, from (17.49), so will $K_R - K_Z$.

Now we show that this entails that each of

$$|f_j(t_0') - f_j(\theta_0)|$$

converges to zero in probability. In fact, since $\left| f_j(\theta_0) - \frac{n_j}{n} \right|$ does so, it will be enough to prove that the same holds for

$$\left| f_j(t_0') - \frac{n_j}{n} \right|. \qquad \ldots \qquad (17.50)$$

Let K_1 be the maximum of K for some fixed z_1. Then $K_R \geqslant K_1$ and

$$K_R - K_M \geqslant K_1 - K_M.$$

Hence $K_1 - K_M$ converges to zero. The maximum K_1 is readily seen to be given by

$$z_j = \frac{n_j(1 - z_1)}{n - n_1}, \qquad j = 2, \ldots, m \qquad \ldots \qquad (17.51)$$

$$K_1 = n_1 \log z_1 + (n - n_1)\{\log(1 - z_1) - \log(n - n_1)\} + \sum_{j=2}^{m} n_j \log n_j. \qquad (17.52)$$

Now z_1 is a double-valued function of K_1, continuous and having its two values equal for $K_1 = K_R$; for K_1 is continuous in z_1 from 0 to 1 (not inclusive), and $\frac{\partial K_1}{\partial z_1}$ changes sign only for $z_1 = n_1/n$, where $K_1 = K_R$. It follows that when $K_R - K_1$ is small, so is $z_1 - n_1/n$. If the other z's are not given by (17.51) $K_R - K$ is smaller still.

A similar argument applies for any j, and hence $\left| z_j - \frac{n_j}{n} \right|$ converges to zero in probability when $K_R - K$ does so. Taking $z_j = f_j(t_0')$ and remembering that in this case K becomes K_Z, we reach (17.50).

Finally, by hypotheses (a) and (b) at least some of the $f_j(\theta)$ have continuous inverse functions expressing θ in terms of the functions f, and hence by taking

$$|f_j(t_0') - f_j(\theta_0)|$$

as small as we please, we may make $t'_0 - \theta_0$ as small as we please. Consequently t' converges to θ_0 in probability and is consistent.

17.24. The reader may find the foregoing proof easier to follow if we express its main points in geometrical terminology.

Consider the m proportions n_j/n as the co-ordinates of a point in a space of m dimensions. The theoretical frequencies $f_j(\theta_0) = 1/m$ define a point, say M, in this space, and the sample point R, corresponding to an observed set of n_j's, may be regarded as varying round the "theoretical" point M. The quantities z are the co-ordinates of any point in the hyperplane $\Sigma(z) = 1$, which contains M and R. (See Fig. 17.1.)

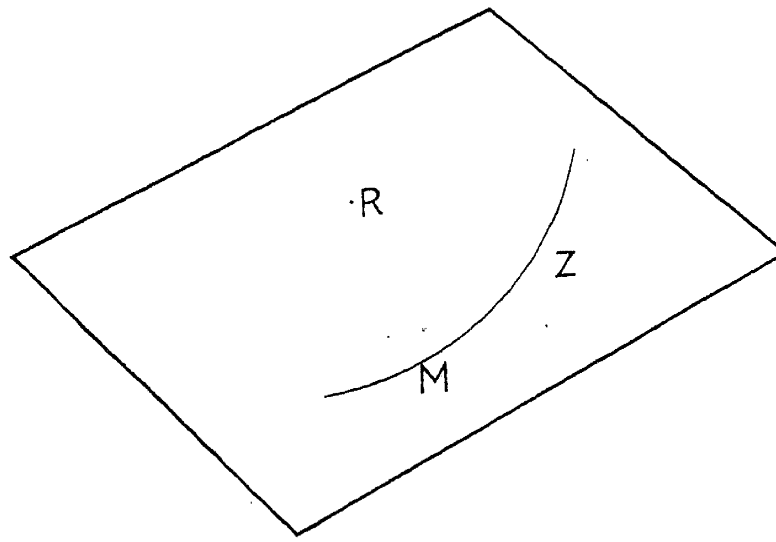


Fig. 17.1.

Now, for any sample point R the maximum likelihood estimator t' assumes a value t'_0 which in general differs from θ_0. This value defines m quantities $f_j(t'_0)$ which determine a point Z. This also lies in the hyperplane since the sum of the frequencies is unity. Thus the points R determine a set of points Z which all lie on the curve defined for variations in θ by

$$z_j = f_j(\theta) \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.53)$$

Since $\theta = \theta_0$ is a possible value of θ, the point M lies on this curve; R in general does not.

What we have shown in analytical form is that the function K, which is the logarithm of a likelihood function defined for any point on the hyperplane, has a maximum at R and a maximum on the curve itself at Z. As the sample size increases, R is as near as we like to M (in the sense of convergence in probability, that is to say, that as high a proportion of points R as we like are as near as we like to M). This involves that Z also is as near as we like to M. This in turn involves that the parameter-value t'_0 corresponding to Z is as close as we like to θ_0 for as high a proportion of the possible points Z as we like, which is our theorem.

17.25. We now prove a second fundamental property of maximum likelihood estimators, namely that they tend to normality for large n. More precisely,

(*a*) If condition (*a*) at the beginning of **17.23** is satisfied; and if (more stringently than condition (*b*) of that section) (*c*) in a θ-interval containing the true value θ_0, $\frac{\partial f}{\partial \theta}$ is continuous in θ for every x, $x^2 \frac{\partial f}{\partial \theta}$ approaches a continuous function of θ as x tends to infinity, and $\frac{\partial f}{\partial \theta}$ does not vanish in some interval,

then the maximum likelihood estimator t tends to normality for large n. The condition as to $x^2 \frac{\partial f}{\partial \theta}$ ensures that in the transformation to finite range $\frac{\partial f}{\partial \theta}$ remains continuous in θ throughout that range.

We recall that if

$$\xi_j = \frac{n_j}{n} - \frac{1}{m}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (17.54)$$

that is, if the ξ's are the deviations of the actual proportional frequencies n_j/n from the "expected" frequencies $1/m$, the distribution of the ξ's in the limit will be normal and their distribution spherically symmetric. Consider again the orthogonal space of the previous section. The sample points are distributed about the point M in a symmetrical form which tends to normality. If we choose a set of orthogonal axes in the hyperplane, the projection of the sample points on any axis is in the limit distributed normally with variance $1/mn$.

In the neighbourhood of M the curve (17.53) approaches its tangent line as n becomes larger, and we therefore have, if s is the distance along the tangent from M,

$$s^2 = (\theta - \theta_0)^2 \sum_{j=1}^{m} \left\{\frac{\partial}{\partial\theta} f_j(\theta_0)\right\}^2, \qquad . \quad . \quad . \quad . \quad (17.55)$$

as follows from (17.53). (The tangent exists in virtue of our hypothesis as to the differential coefficients of f in θ.)

Now consider the point Z on the curve corresponding to the sample point R. We know that at Z the function

$$K = \Sigma n_j \log\left(z_j + \frac{1}{m}\right), \qquad . \quad . \quad . \quad . \quad . \quad (17.56)$$

where we now measure z from M, is a maximum for variations in z such that Z lies on the curve. R is determined by finding the hypersurface (17.56) tangent to the hyperplane $\Sigma(z_j) = 0$, for at that point $\partial K/\partial z_j$ is zero. We know that the co-ordinates of this point are $z_j = n_j/n - 1/m$ and that R is the point of tangency. K_R as defined in **17.23** is the value of K at R, and K_Z is that at Z. We then have, by Taylor's theorem,

$$K_Z = K_R + \sum_j \left(\frac{\partial K}{\partial z_j}\right)_R \delta z_j + \tfrac{1}{2}\sum_{j,k}\left(\frac{\partial^2 K}{\partial z_j\,\partial z_k}\right)_R \delta z_j\,\delta z_k \qquad . \quad . \quad (17.57)$$

to the second order of small quantities in δz. From (17.56) we see that

$$\frac{\partial K}{\partial z_j} = n \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.58)$$

$$\left.\begin{aligned}\frac{\partial^2 K}{\partial z_j\,\partial z_k} &= 0, & j \neq k\\ &= -\frac{n^2}{n_j}, & j = k\end{aligned}\right\} \qquad . \quad . \quad . \quad . \quad (17.59)$$

Hence

$$K_Z = K_R + n\,\Sigma(\delta z_j) - \frac{n^2}{2}\Sigma\frac{(\delta z_j)^2}{n_j} \qquad . \quad . \quad . \quad . \quad (17.60)$$

Now $\Sigma(\delta z_j) = 0$, for the variation takes place in the hyperplane. Hence, for given R, Z is the point for which $\Sigma\frac{(\delta z_j)^2}{n_j}$ is a minimum. As n tends to infinity the n_j's tend to equality, and hence Z is the point on the curve which is nearest to R. Thus R is, in the limit, projected orthogonally on to the curve, that is to say, in the limit, on the tangent line.

Now we know that these points are distributed normally with variance $1/mn$ and

this proves the theorem. We may also evaluate the variance of the maximum likelihood estimator; for

$$\operatorname{var} t' = \frac{\operatorname{var} s}{\sum_{j=1}^{m}\left\{\frac{\partial}{\partial\theta}f_j(\theta)\right\}^2}$$

$$= \frac{1}{mn\,\Sigma\left\{\frac{\partial}{\partial\theta}f_j(\theta)\right\}^2} \quad . \quad . \quad . \quad . \quad . \quad (17.61)$$

and since t' approaches t for fine grouping we have also, remembering that $1/m = f_j(\theta_0)$,

$$\frac{1}{\operatorname{var} t} = n\int_{-\infty}^{\infty}\left(\frac{\partial f}{\partial\theta}\right)^2\frac{dx}{f}$$

$$= n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial\theta}\right)^2 f\,dx, \quad . \quad . \quad . \quad . \quad . \quad (17.62)$$

where θ is to be put equal to θ_0 on the right.

It may be remarked that condition (c) at the beginning of the section prevents the vanishing of $\frac{\partial f}{\partial\theta}$ which might render the expression (17.61) nugatory.

17.26. We have, then, under the afore-mentioned conditions,

$$\frac{1}{\operatorname{var} t} = n\,E\left(\frac{\partial \log f}{\partial\theta}\right)^2.$$

If the range is independent of θ, or if f and $\frac{\partial f}{\partial\theta}$ vanish at any extremity of the range which depends on θ, we have the alternative form—

$$\frac{1}{\operatorname{var} t} = -\,n\,E\left(\frac{\partial^2 \log f}{\partial\theta^2}\right). \quad . \quad . \quad . \quad . \quad . \quad (17.63)$$

In fact, since $\int_a^b f\,dx = 1$ where a, b are the limits of the range and may contain θ, we have *

$$0 = \frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b f\frac{\partial \log f}{\partial\theta}\,dx + f(b,\theta)\frac{\partial b}{\partial\theta} - f(a,\theta)\frac{\partial a}{\partial\theta}$$

$$= \int_a^b f\left(\frac{\partial \log f}{\partial\theta}\right)dx.$$

Differentiating again, we have

$$0 = \int_a^b\left(\frac{\partial \log f}{\partial\theta}\right)^2 f\,dx + \int_a^b\left(\frac{\partial^2 \log f}{\partial\theta^2}\right)f\,dx + \left(f\frac{\partial \log f}{\partial\theta}\right)\frac{\partial b}{\partial\theta} - \left(f\frac{\partial \log f}{\partial\theta}\right)\frac{\partial a}{\partial\theta}. \quad . \quad (17.64)$$

Again, if the range is independent of θ or if $\left(\frac{\partial f}{\partial\theta}\right)$ vanishes at the extremity, the last two

* The operation of differentiating under the integral sign requires certain conditions as to uniform convergence, even when the limits are independent of θ. To avoid prolixity we shall always assume that the conditions hold unless the contrary is stated. The point gives rise to no statistical difficulty but is troublesome when one is aiming at complete mathematical rigour.

terms on the right in (17.64) are zero, and we have (reverting to our usual convention as to limits)

$$\int_{-\infty}^{\infty} \frac{\partial^2 (\log f)}{\partial \theta^2} f\, dx = -\int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f\, dx$$

and the result follows from (17.62).

17.27. We now prove a third fundamental property concerning the efficiency of maximum likelihood estimates.

If t be any estimator of θ, the range of $f(x, \theta)$ is independent of θ, and in large samples t is distributed normally about mean θ_0 (the true value of θ) with variance v; then

$$\frac{1}{nv} \text{ cannot exceed } \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f\, dx, \text{ with } \theta = \theta_0;$$

and hence, if a maximum likelihood estimator exists, it is most-efficient in the class of such estimators.

By hypothesis, we have in the limit for the frequency function of t,

$$\Phi = \frac{1}{\sqrt{(2\pi v)}} \exp\left\{-\frac{(t-\theta)^2}{2v}\right\} \quad . \quad . \quad . \quad . \quad (17.65)$$

and hence

$$\frac{\partial^2 \log \Phi}{\partial \theta^2} = -\frac{1}{v}, \quad . \quad . \quad . \quad . \quad . \quad (17.66)$$

where, for convenience, we drop the suffix of θ until the end of the proof. We then have

$$\begin{aligned} \frac{1}{v} &= \int_{-\infty}^{\infty} -\frac{\partial^2 \log \Phi}{\partial \theta^2} \Phi\, dt \\ &= \int_{-\infty}^{\infty} \frac{1}{\Phi}\left(\frac{\partial \Phi}{\partial \theta}\right)^2 dt. \quad . \quad . \quad . \quad . \quad . \quad (17.67) \end{aligned}$$

Now consider

$$u = \frac{\partial}{\partial \theta} (\log L) \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.68)$$

as a random variable over the possible values $x_1 \ldots x_n$ conditioned by $t =$ constant. Since the frequency of u is L, we have

$$\text{var } u = \frac{\Sigma (Lu^2)}{\Sigma (L)} - \frac{\{\Sigma (Lu)\}^2}{\{\Sigma (L)\}^2} \quad . \quad . \quad . \quad . \quad (17.69)$$

with summation (or integration) over the range of x's. Now Φ is the frequency of all samples having a constant t, and hence

$$\Phi = \Sigma (L).$$

Hence

$$\begin{aligned} \text{var } u &= \frac{\Sigma (Lu^2)}{\Phi} - \frac{\{\Sigma (Lu)\}^2}{\Phi^2} \\ &= \frac{1}{\Phi} \Sigma \left\{\frac{1}{L}\left(\frac{\partial L}{\partial \theta}\right)^2\right\} - \frac{1}{\Phi^2}\left\{\Sigma\left(\frac{\partial L}{\partial \theta}\right)\right\}^2 \quad . \quad . \quad . \quad (17.70) \end{aligned}$$

Now var u cannot be negative and Φ is not negative, and hence

$$\Sigma \left\{\frac{1}{L}\left(\frac{\partial L}{\partial \theta}\right)^2\right\} - \frac{1}{\Phi}\left\{\Sigma\left(\frac{\partial L}{\partial \theta}\right)\right\}^2 \geqslant 0. \quad . \quad . \quad . \quad (17.71)$$

But
$$\Sigma\left(\frac{\partial L}{\partial \theta}\right) = \frac{\partial}{\partial \theta}(\Sigma L) = \frac{\partial \Phi}{\partial \theta},$$
and hence, substituting in (17.71) and integrating over all t, we have
$$\int dt\,\Sigma\left\{\frac{1}{L}\left(\frac{\partial L}{\partial \theta}\right)^2\right\} \geqslant \int dt\,\frac{1}{\Phi}\left(\frac{\partial \Phi}{\partial \theta}\right)^2 = \frac{1}{v}. \quad . \quad . \quad . \quad (17.72)$$
Now Σ is carried out over all x for constant t and the integration over all t, so that the two summations together are equivalent to summation over the x's without restriction. Hence
$$\frac{1}{v} \leqslant \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \frac{1}{L}\left(\frac{\partial L}{\partial \theta}\right)^2 dx_1 \ldots dx_n$$
$$\leqslant \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} L\left(\frac{\partial \log L}{\partial \theta}\right)^2 dx_1 \ldots dx_n$$
$$\leqslant n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2 f\,dx$$
which establishes the result, since the expression on the right is the reciprocal of the variance of the maximum likelihood estimator, if it exists.

17.28. The fourth fundamental theorem of maximum likelihood estimators is as follows :—

If a sufficient estimator exists, it is a function of the maximum likelihood estimator.

In fact, the likelihood can then be put in the form
$$L = L_1(t, \theta)\, L_2(x_1 \ldots x_n),$$
where L_2 does not contain θ. Hence
$$\frac{\partial}{\partial \theta}\log L = \frac{\partial}{\partial \theta}\log L_1$$
$$= \psi(\theta, t), \text{ a function of } \theta \text{ and } t \text{ only.} \quad . \quad (17.73)$$
Hence, for fixed t, $\frac{\partial}{\partial \theta}\log L$ is constant, and it follows from the previous section that the variance of t is equal to the variance of a most-efficient estimator (for var u is then zero for fixed t and the inequality (17.72) becomes an equality). Hence the sufficient estimator is most-efficient, confirming the result of **17.18.**

It follows from (17.73) that the maximum likelihood estimator is given by
$$\psi(\theta, t) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.74)$$
which proves the theorem.

Conversely, if t is such that (17.73) is true, it must be sufficient ; for then we have
$$\log L = C + \int \psi(\theta, t)\, d\theta,$$
where C does not depend on θ and the likelihood is of the requisite form.

Example 17.7

Consider the estimation of the parameter m in the population
$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}dx, \qquad -\infty \leqslant x \leqslant \infty$$

where σ is known. The frequency function is easily seen to obey the conditions relating to maximum likelihood estimators. We have

$$\log L = -n \log \sigma\sqrt{(2\pi)} - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - m)^2,$$

and hence the maximum likelihood estimator is the root of

$$\frac{\partial}{\partial m}\log L = \frac{1}{\sigma^2}\Sigma(x - m) = 0,$$

giving

$$\hat{m} = \frac{1}{n}\Sigma(x) = \bar{x}.$$

It is frequently convenient to denote the estimator of a parameter by writing a circumflex accent over it in this way.

In this case the sample mean is the maximum likelihood estimator. It is therefore most-efficient and no other estimator can have a smaller variance in the limit. For the variance we have, from (17.63),

$$\begin{aligned}\frac{1}{\operatorname{var}\hat{m}} &= -n\int_{-\infty}^{\infty}\left(\frac{\partial^2 \log f}{\partial\theta^2}\right)_{\theta=m} f\,dx\\ &= n\int_{-\infty}^{\infty}\frac{1}{\sigma^2}f\,dx\\ &= \frac{n}{\sigma^2},\end{aligned}$$

giving the familiar result—

$$\operatorname{var}\bar{x} = \frac{\sigma^2}{n}.$$

This, as it happens, is true for any n. The estimator is also sufficient, for

$$\begin{aligned}\frac{\partial}{\partial m}\log L &= \frac{1}{\sigma^2}(n\bar{x} - nm)\\ &= \text{a function of } m \text{ and } \bar{x} \text{ only.}\end{aligned}$$

The condition that σ^2 is known is to be noted. Complications arise when two parameters are estimated simultaneously, as we shall see presently.

Example 17.8

Consider the estimation of θ in the Type III distribution

$$dF = \frac{x^{p-1}e^{-x/\theta}}{\Gamma(p)\,\theta^p}dx, \qquad 0 \leqslant x \leqslant \infty$$

where p is known.

We have

$$\log f = (p-1)\log x - \frac{x}{\theta} - \log\Gamma(p) - p\log\theta$$

and hence, dropping terms independent of θ,

$$\log L = -\frac{1}{\theta}\Sigma(x) - np\log\theta.$$

The equation of maximum likelihood is then

$$\frac{1}{\theta^2}\Sigma(x) - \frac{np}{\theta} = 0,$$

giving

$$\hat{\theta} = \frac{\Sigma(x)}{np} = \frac{\bar{x}}{p}.$$

The variance is given, by (17.63), as

$$\frac{1}{\text{var}\,\hat{\theta}} = -n\int_{-\infty}^{\infty}\left(-\frac{2x}{\theta^3} + \frac{p}{\theta^2}\right)f\,dx$$

$$= -n\left\{\frac{p}{\theta^2} - \frac{2p}{\theta^2}\right\};$$

$$\text{var}\,\hat{\theta} = \frac{\theta^2}{np},$$

where θ is the true value of the parameter. We could also have obtained this result directly (and again it happens to be true for all n). From Example 10.11 (vol. I, p. 244) we have for the distribution of $\bar{x}/p = \hat{\theta}$,

$$dF = n^{np}\left(\frac{p}{\theta}\right)^{np}\frac{\hat{\theta}^{np-1}\exp\left(-\frac{np\hat{\theta}}{\theta}\right)}{\Gamma(np)}\,d\hat{\theta},$$

from which the first two moments about the origin are

$$\mu_1' = \theta, \qquad \mu_2' = \frac{np+1}{np}\theta^2,$$

giving

$$\text{var}\,\hat{\theta} = \mu_2 = \frac{\theta^2}{np}.$$

We note that the likelihood function may be put in the form

$$\log L = (p-1)\,\Sigma\log x - n\log\Gamma(p) - \frac{np\hat{\theta}}{\theta} - np\log\theta,$$

from which it is evident that $\hat{\theta}$ is sufficient.

Example 17.9

Consider the estimation of the parameter λ in the Poisson distribution whose general term is $e^{-\lambda}\dfrac{\lambda^x}{x!}$.

In this case the likelihood function is discontinuous and we have

$$L = \frac{e^{-n\lambda}\lambda^{\Sigma(x)}}{x_1!\;.\;.\;.\;x_n!}.$$

Hence

$$\frac{\partial}{\partial\lambda}\log L = -n + \frac{n\bar{x}}{\lambda},$$

giving $\hat{\lambda} = \bar{x}$, the sample mean.

For the variance we have

$$\frac{1}{\operatorname{var}\hat{\lambda}} = n \sum_{x=0}^{\infty} \left(\frac{x}{\lambda^2} e^{-\lambda} \frac{\lambda^x}{x\,!}\right)$$

$$= \frac{n}{\lambda}$$

$$\operatorname{var}\hat{\lambda} = \frac{\lambda}{n}, \text{ a familiar result.}$$

It is easy to see in this case also that $\hat{\lambda}$ is sufficient.

Example 17.10

What is the most general form of distribution, differentiable in θ, for which the sample-mean is the maximum likelihood estimator ?

We are given that a solution of

$$\frac{\partial}{\partial\theta} \log L = \Sigma\left(\frac{1}{f}\frac{\partial f}{\partial\theta}\right) = 0$$

is

$$\theta = \frac{1}{n}\Sigma(x)$$

or

$$\Sigma(x - \theta) = 0.$$

This is true for all x and θ, and hence

$$\frac{1}{f}\frac{\partial f}{\partial\theta} = (x - \theta)\, K,$$

where K is independent of x but may be dependent on θ, say equal to $\frac{\partial^2\psi}{\partial\theta^2}$. Then, integrating,

$$\log f = \int d\theta\,(x - \theta)\frac{\partial^2\psi}{\partial\theta^2}$$

$$= (x - \theta)\frac{\partial\psi}{\partial\theta} + \psi + \zeta(x),$$

where $\zeta(x)$ is an arbitrary function of x. Hence

$$f = k \exp\left\{(x - \theta)\frac{\partial\psi}{\partial\theta} + \psi(\theta) + \zeta(x)\right\},$$

which is the most general form of f.

If $\qquad \psi(\theta) = \frac{1}{2}\theta^2, \quad \zeta(x) = -\frac{1}{2}x^2$

the form becomes the normal distribution

$$f = k \exp\{-\tfrac{1}{2}(x - \theta)^2\}.$$

Successive Approximations to Efficient Estimators

17.29. In the examples we have just given, the solution of the maximum likelihood equation was carried out without difficulty. It frequently happens, however, that the equation is by no means so easy to solve explicitly, though it can sometimes be solved

for particular values of x by iterative methods. Another possibility is to compute an inefficient estimator and correct it by an extra term, which can be obtained as follows :—

Let t' be an inefficient estimator and t a most-efficient estimator. Let

$$\delta = t' - t.$$

Then

$$\operatorname{var} \delta = \operatorname{var} t' + \operatorname{var} t - 2 \operatorname{cov} (t', t). \qquad (17.75)$$

Remembering that if E is the efficiency of t',

$$\operatorname{var} t = E \operatorname{var} t'$$

$$\frac{\operatorname{cov} (t', t)}{(\operatorname{var} t \ \operatorname{var} t')^{\frac{1}{2}}} = \sqrt{E} \qquad \text{(see (17.39))};$$

we have

$$\operatorname{var} \delta = \frac{1-E}{E} \operatorname{var} t. \qquad (17.76)$$

If then t' is "nearly" efficient, that is, if $1 - E$ is small, the average value of $\delta = t' - t$ will be small.

If the maximum likelihood equation is

$$\left(\frac{\partial L}{\partial \theta}\right)_{\theta=t} = 0,$$

consider

$$t'' = t' + \operatorname{var} t \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t'} \qquad (17.77)$$

We have

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t'} = \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t} + (t'-t)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta=t} + \text{terms of higher order}$$

$$= (t'-t)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta=t} \qquad (17.78)$$

For large n, approximately

$$-\frac{1}{\operatorname{var} t} = \left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta=t}$$

and hence, approximately,

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t'} = \frac{t-t'}{\operatorname{var} t}.$$

Hence

$$\begin{aligned} t'' &= t' + \operatorname{var} t \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t'} \\ &= t' + t - t' \\ &= t, \end{aligned}$$

and t'' is an efficient estimator to a better order of approximation. This process may be repeated and, rather like Newton's successive approximation to the roots of an equation, may be expected to improve the efficiency of an estimator.

Example 17.11

Suppose we have to estimate θ, the parameter in the Cauchy population

$$dF = \frac{1}{\pi} \frac{dx}{1+(x-\theta)^2}, \qquad -\infty \leqslant x \leqslant \infty.$$

We have already seen that the sample-mean is not a satisfactory estimate and that for large samples the median is consistent and has variance $\pi^2/4n$.

The equation of maximum likelihood gives

$$-\frac{\partial \log L}{\partial \theta} = \Sigma \left\{ \frac{2(x-\theta)}{1+(x-\theta)^2} \right\} = 0.$$

This is a $(2n-1)$-ic in θ and correspondingly difficult to solve. We may, however, find the variance of the solution $\hat{\theta}$ from (17.63). We have

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\partial^2 \log f}{\partial \theta^2} f\,dx &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{2(x-\theta)^2-2}{\{1+(x-\theta)^2\}^3}\,dx \\ &= \frac{4}{\pi}\int_0^{\infty} \frac{(x^2-1)\,dx}{(1+x^2)^3} \\ &= -\frac{1}{2}.\end{aligned}$$

Hence

$$\text{var}\,\hat{\theta} = \frac{2}{n}.$$

The median, therefore, has an efficiency of $\frac{8}{\pi^2} = 0{\cdot}8$, and we expect that

$$\begin{aligned}t'' &= t' + \text{var}\,\hat{\theta}\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t'} \\ &= t' - \frac{4}{n}\Sigma\left\{\frac{x-t'}{1+(x-t')^2}\right\},\end{aligned}$$

where t' denotes the median, will be an improved estimator.

Most General Form of Distributions possessing Sufficient Estimators

17.30. If t is sufficient for θ we have

$$\frac{\partial \log L}{\partial \theta} = K(t, \theta), \qquad (17.79)$$

where K is some function of t and θ. Regarding this as an equation in t we see that it remains true for any particular value of θ, say zero. It is then evident that t must be expressible in the form

$$t = M\left\{\sum_{j=1}^{n} k(x_j)\right\}, \qquad (17.80)$$

where M and k are arbitrary functions. If $w = \Sigma\, k(x)$ then K is a function of θ and w only, say $N(t, w)$. We have then

$$\frac{\partial^2 \log L}{\partial \theta\, \partial x_j} = \frac{\partial N}{\partial w}\frac{\partial w}{\partial x_j}. \qquad (17.81)$$

Now the left-hand side is a function of θ and x_j only and w is a function of x_j only. Hence $\frac{\partial N}{\partial w}$ is a function of θ and x_j only. But it must be symmetrical in the x's and hence is a function of θ only. Hence, integrating with respect to w, we have

$$N(t, w) = wp(\theta) + q(\theta),$$

where p and q are arbitrary functions of θ. Thus

$$\frac{\partial}{\partial\theta}(\log L) = \frac{\partial}{\partial\theta}\sum_j \{\log f(x_j, \theta)\} = p(\theta)\sum_j k(x_j) + q(\theta) \quad . \quad . \quad (17.82)$$

whence
$$\frac{\partial}{\partial\theta}\log f(x, \theta) = p(\theta)\,k(x) + \frac{1}{n}q(\theta),$$

giving
$$f(x, \theta) = \exp\{p(\theta)\,k(x) + q(\theta) + r(x)\}, \quad . \quad . \quad . \quad (17.83)$$

where we still write p and q for the integrated functions.

The expression may also be written

$$f(x, \theta) = Q(\theta)\,R(x)\exp\{p(\theta)\,k(x)\} \quad . \quad . \quad . \quad (17.84)$$

or, if we simplify the specification of the distribution by writing θ instead of $p(\theta)$,

$$f(x) = Q(\theta)\,R(x)\exp\{\theta\,k(x)\}. \quad . \quad . \quad . \quad . \quad (17.85)$$

It will be found that if (17.85) holds, the likelihood function is of the required form for the existence of a sufficient estimator, so that the equation is sufficient as well as necessary.

Distribution of Sufficient Estimators

17.31. It is remarkable that the distribution of a sufficient estimator can be obtained directly from the likelihood function. From (17.85) we have

$$\log L = n\log Q + \Sigma\log R(x) + \theta\,\Sigma\,k(x)$$

giving, for the maximum likelihood estimator,

$$\frac{n}{Q}\frac{\partial Q}{\partial\theta} + \Sigma\,k(x) = 0. \quad . \quad . \quad . \quad . \quad . \quad (17.86)$$

Now, for the characteristic function $\phi(\alpha)$ of $w\ (=\Sigma\,k(x))$ we have—

$$\begin{aligned}\phi(\alpha) &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{iw\alpha} f(x_1, \theta)\,dx_1 \ldots f(x_n, \theta)\,dx_n \\ &= \left\{\int_{-\infty}^{\infty} e^{ik(x)\alpha} f\,dx\right\}^n \\ &= \left\{\int_{-\infty}^{\infty} Q(\theta)\,R(x)\,e^{(i\alpha+\theta)\,k(x)}\,dx\right\}^n \\ &= \left\{\frac{Q(\theta)}{Q(\theta + i\alpha)}\right\}^n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.87)\end{aligned}$$

Hence the frequency function of w, if existent, is

$$f(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha w}\left\{\frac{Q(\theta)}{Q(\theta + i\alpha)}\right\}^n d\alpha.$$

Now from (17.86),

$$\begin{aligned} w &= -\left(\frac{n}{Q}\frac{\partial Q}{\partial\theta}\right)_{\theta=t} \\ &= n\,S(t), \text{ say,}\end{aligned}$$

and hence the frequency function of the estimator t is

$$f(t) = \frac{n}{2\pi}\left(\frac{\partial S}{\partial t}\right)\int_{-\infty}^{\infty} e^{-i\alpha nS(t)}\left\{\frac{Q(t)}{Q(t + i\alpha)}\right\}^n d\alpha. \quad . \quad . \quad (17.88)$$

Example 17.12

The normal distribution with unit variance may be put in the form

$$f = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}\theta^2} e^{x\theta}.$$

Comparing this with (17.85), we see that if

$$Q(\theta) = e^{-\frac{1}{2}\theta^2}$$

$$R(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

$$k(x) = x$$

the condition for a sufficient estimator is satisfied. That this is (as we already know) the mean $\bar{x}$ may be confirmed from (17.88). We have

$$S(\theta) = -\frac{\partial}{\partial\theta} \log Q = \theta;$$

and hence for the frequency function of the estimator $\bar{x}$,

$$\frac{n}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha n\bar{x}} \left\{\frac{e^{-\frac{1}{2}\bar{x}^2}}{e^{-\frac{1}{2}(x+i\alpha)^2}}\right\}^n d\alpha$$

$$= \frac{n}{2\pi}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}n\alpha^2 - i\alpha n(\bar{x}-\theta)\}\, d\alpha$$

$$= \sqrt{\frac{n}{2\pi}} \exp\{-\tfrac{1}{2}n(\bar{x}-\theta)^2\}.$$

Example 17.13

The Type III distribution considered in Example 17.8 may be put in the slightly different form

$$dF = \frac{\gamma^p}{\Gamma(p)} x^{p-1} e^{-\gamma x}\, dx, \qquad 0 \leqslant x \leqslant \infty.$$

Regarding p as known and considering γ as the parameter under estimate, we see that a sufficient estimator exists, because we may write

$$Q(\gamma) = \frac{\gamma^p}{\Gamma(p)}$$

$$R(x) = x^{p-1}$$

$$k(x) = x,$$

which throws the distribution into the form (17.85). We have found the estimator and its distribution in Example 17.8.

On the other hand, suppose that γ is known and we wish to estimate p. Writing

$$Q(p) = \frac{\gamma^p}{\Gamma(p)}$$

$$R(x) = e^{-\gamma x - \log x}$$

$$k(x) = \log x$$

we see that a sufficient estimator for p also exists. It is the solution of

$$-\frac{d}{dp}\log\Gamma(p) + \log\gamma + \frac{1}{n}\Sigma\log x = 0,$$

which does not permit of expression of p as a simple function of the x's. The sampling distribution is not expressible in a simple form.

Example 17.14

Consider again the Cauchy distribution

$$dF = \frac{1}{\pi} \frac{dx}{1 + (x - \theta)^2}, \qquad -\infty \leqslant x \leqslant \infty.$$

Evidently this cannot be thrown into the form (17.85) and hence no sufficient estimator exists. We have already found (Example 17.11) that there is an efficient estimator. For finite n no single estimator can contain all that the sample can tell us about θ.

Sufficient Estimators when the Range depends on the Parameter

17.32. One of the conditions of the theorem of **17.23** and that of **17.27** is that the range should be independent of θ. In the contrary case our results, particularly for sufficient estimators, require reconsideration.

Suppose the range of the frequency function is from θ to b, where b is fixed. If there is a sufficient estimator for θ, say t, the distribution of t and any other estimator is independent of θ. Take x_1, the lowest value of the sample, as such other estimator. Then if t is fixed the distribution of x_1 is independent of θ, which is clearly impossible unless in fixing t we also fix x_1, that is to say, t is a function of x_1. Thus if a sufficient estimator exists it must be a function of x_1.

Similarly if the range is from a to θ, a sufficient estimator for θ must be a function of the largest sample member.

17.33. If x_1 or some function of it is sufficient for θ, the lower extremity of the range, and x_1 is fixed, the probability that any particular sample value x is greater than x_1 is proportional to $f(x, \theta)$. This must be independent of θ, since x_1 is sufficient, and hence so is $f(x, \theta)/f(x_1, \theta)$. Thus

$$f(x, \theta) = \frac{g(x)}{h(\theta)}, \qquad (17.89)$$

and this is the most general form admitting a sufficient estimator.

It remains true in such circumstances that the smallest member of the sample is a maximum likelihood estimator. For the likelihood is

$$L = \frac{g(x_1) \ . \ . \ . \ g(x_n)}{\{h(\theta)\}^n},$$

which is clearly a maximum when $h(\theta)$ is a minimum. Now since the total frequency is unity we have, from (17.89),

$$h(\theta) = \int_\theta^b g(x)\, dx. \qquad (17.90)$$

θ cannot be greater than x_1, for then such a sample value could not appear. The value which minimises $h(\theta)$ is seen from (17.90) to be that which minimises the range, i.e. x_1.

17.34. When both extremes of the range, a and b, depend on θ, some further modification is necessary. Suppose that a is equal to θ and that $b(\theta)$ is some strictly decreasing

function of θ. Let X_n be the value such that $b(X_n) = x_n$, the greatest member of the sample, and let t be the smaller of x_1 and X_n. Then of the inequalities

$$t \leqslant x_1, \qquad b(t) \geqslant x_n \qquad (17.91)$$

one at least is true. But the first equality implies that $t \geqslant \theta$ and the second that $b(t) \leqslant b(\theta)$, and either of these two implies the other. Hence both inequalities in (17.91) are true, and

$$\theta \leqslant t \leqslant x_1 \leqslant x_n \leqslant b(t) \leqslant b(\theta). \qquad (17.92)$$

Samples with fixed t then lie in a fixed range, and hence t is sufficient if the frequency function is of the form (17.89). It would seem that this remains the most general form of frequency function admitting a sufficient estimator when both extremes of the range depend on θ.

Example 17.15

Consider the rectangular distribution

$$dF = \frac{dx}{2\theta}, \qquad -\theta \leqslant x \leqslant \theta.$$

If we take the ordinary likelihood equation we get

$$\frac{\partial}{\partial\theta}\log L = -\frac{\partial}{\partial\theta}n\log(2\theta) = -\frac{n}{\theta}.$$

For this to vanish θ must tend to infinity, an obviously nugatory result. In accordance with the above discussion we should take as our estimate of θ the smaller of x_1 and $-x_n$, and this is obviously sufficient, for nothing in the sample can tell us more about the terminals of the range than its most extreme members.

Intrinsic Accuracy

17.35. If the sampling distribution of an estimator t is

$$dF = \Phi(t, \theta)\,dt \qquad (17.93)$$

we define the *accuracy* of t as

$$I' = \int_{-\infty}^{\infty}\left(\frac{\partial\Phi}{\partial\theta}\right)^2\frac{1}{\Phi}\,dt$$
$$= E\left(\frac{\partial\log\Phi}{\partial\theta}\right)^2. \qquad (17.94)$$

It is evidently essentially a positive quantity. We assume, unless the contrary is stated, that the range is independent of θ.

I' is the quantity we have already encountered in (17.67) as the reciprocal of the variance of t when it tends to normality in large samples. As in **17.27**, we have

$$I' \leqslant n\int_{-\infty}^{\infty}\left(\frac{\partial\log f}{\partial\theta}\right)^2 f\,dx \qquad (17.95)$$

$\leqslant nI$, say, where

$$I = \int_{-\infty}^{\infty}\left(\frac{\partial\log f}{\partial\theta}\right)^2 f\,dx. \qquad (17.96)$$

Now I is independent of the estimator t and we may call it the *intrinsic accuracy* of the distribution f in regard to θ. It is intrinsic because it depends only on f. It may

be termed accuracy because it provides, for large samples at least, a minimum to the variance of possible estimators of θ. We know from **17.25** that under certain conditions the maximum likelihood estimator attains this minimum for large samples.

17.36. We may now extend the definition of efficiency of an estimator to the case of small samples. In fact, the efficiency is the ratio of the accuracy of an estimator to the intrinsic accuracy of the distribution for the parameter under estimate. This is easily seen to apply to the case of large samples for which efficiency was defined in **17.12**, and may be applied to finite samples or non-normal sampling variation. For such cases, however, it is conceivable that the efficiency might exceed unity. A proof that this is not so when the range is independent of θ is suggested in Exercise 17.12.

17.37. If the range is independent of θ we have

$$E\left(\frac{\partial \log f}{\partial \theta}\right) = \int \frac{\partial f}{\partial \theta}\, dx = \frac{\partial}{\partial \theta}\int f\, dx = 0$$

and hence the following three expressions for the intrinsic accuracy are equivalent:

$$\left.\begin{array}{r} E\left(\dfrac{\partial \log f}{\partial \theta}\right)^2 \\ -\,E\left(\dfrac{\partial^2 \log f}{\partial \theta^2}\right) \\ \operatorname{var}\left(\dfrac{\partial \log f}{\partial \theta}\right) \end{array}\right\} \qquad (17.97)$$

This equivalence holds if f is zero at the extremes of the range. For we then have

$$\begin{aligned} 0 = \frac{\partial}{\partial \theta}\int_a^b f\, dx &= \int_a^b \frac{\partial f}{\partial \theta}\, dx - f(a, \theta)\frac{\partial a}{\partial \theta} + f(b, \theta)\frac{\partial b}{\partial \theta} \\ &= \int_a^b \frac{\partial f}{\partial \theta} dx. \end{aligned}$$

But if f is not zero at the extremes the equivalence may break down. (Cf. Exercises 17.9 and 17.11.)

Amount of Information

17.38. The quantity nI has been called the *amount of information* about θ in the sample of n, and I may be called the amount of information per member of the sample. The use of "information" in this specialised sense has not been universally accepted, but some of the properties of I are such as we should require of any measure of information.

(*a*) If the parent does not contain θ, $I = 0$ so that no sample can tell us anything about θ, which must obviously be so.

(*b*) Since sufficient estimators contain all the relevant information in the sample we expect their accuracy to be nI, and conversely. That this is so may be seen as in **17.27** and **17.28**. In fact, if t is such that the equality in (17.72) holds, var $u = 0$ and for fixed t, $\dfrac{\partial \log L}{\partial \theta}$ is constant, irrespective of the form of distribution of t. Log L is then of the type required for sufficiency.

(*c*) The sum of the amounts of information in two independent sample-members is the amount of information in the pair taken together. For if their joint distribution is

$$dF = f_1(x, \theta)\, dx\, f_2(y, \theta)\, dy,$$

we have for the intrinsic accuracy

$$\begin{aligned} &-\iint \frac{\partial^2 \log f_1 f_2}{\partial \theta^2} f_1 f_2\, dx\, dy \\ &= -\iint \frac{\partial^2 \log f_1}{\partial \theta^2} f_1 f_2\, dx\, dy - \iint \frac{\partial^2 \log f_2}{\partial \theta^2} f_1 f_2\, dx\, dy \\ &= -\int \frac{\partial^2 \log f_1}{\partial \theta^2} f_1\, dx - \int \frac{\partial^2 \log f_2}{\partial \theta^2} f_2\, dy, \end{aligned} \qquad (17.98)$$

which is the property stated.

Loss of Accuracy

17.39. Where no sufficient estimator exists, it follows from (*b*) of the previous paragraph that no estimator for finite n can contain all the information in the sample. In so far as any particular estimator falls short of the ideal we may be said to lose information by using it. No estimator can avoid losing something, although of course some may lose less than others.

Presumably the loss will be greater for large samples than for small ones, and will be least for maximum likelihood estimators. We may calculate the loss in this case. If t is the maximum likelihood estimator of θ, we have, to a first approximation,

$$\frac{\partial \log L}{\partial \theta} = (\theta - t)\frac{\partial^2 \log L}{\partial \theta^2}. \qquad (17.99)$$

The variance of $\dfrac{\partial \log L}{\partial \theta}$ in samples for which t is constant is thus the variance of $\dfrac{\partial^2 \log L}{\partial \theta^2}$ within the set multiplied by $(t - \theta)^2$. Now the total loss of information, from **17.27**, is seen to be $\operatorname{var} u = \operatorname{var}\left(\dfrac{\partial \log L}{\partial \theta}\right)$, and hence is equal to the variance of t multiplied by the total variance of $\dfrac{\partial^2 \log L}{\partial \theta^2}$ within sets for which t is constant. This we now evaluate.

Suppose the distribution is grouped so that the "expected" frequency in the jth group is m_j. The likelihood is then proportional to $m_1^{n_1} m_2^{n_2} \ldots$ and apart from constants independent of θ we have

$$\log L = \sum_j n_j \log m_j \qquad (17.100)$$

$$\frac{\partial \log L}{\partial \theta} = \Sigma \frac{m'}{m} n, \text{ where } m' = \frac{\partial m}{\partial \theta} \qquad (17.101)$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \Sigma\left\{\left(\frac{m''}{m} - \frac{m'^2}{m^2}\right) n\right\}. \qquad (17.102)$$

We have at once

$$\begin{aligned} \frac{1}{\operatorname{var} t} &= -E\,\Sigma\left\{\left(\frac{m''}{m} - \frac{m'^2}{m^2}\right) n\right\} = -E\,\Sigma\left\{m'' - \frac{m'^2}{m}\right\} \\ &= \Sigma\left(\frac{m'^2}{m}\right). \end{aligned} \qquad (17.103)$$

We shall find it most convenient to regard the n's as distributed over the groups first of all without restriction and then subject to two linear constraints expressed by $\Sigma(n_j) = n$ and $\dfrac{\partial \log L}{\partial \theta} = \Sigma\left(\dfrac{m'}{m} n\right) =$ constant. From this viewpoint the n's may be regarded as distributed in the Poisson form with mean and variance m (not the binomial because we are not introducing the restriction that the samples should be of fixed size, except as a constraint).

Now if $\Sigma(k_j n_j)$ is a linear function of the n's subject to a linear constraint $\Sigma(\alpha_j n_j) = p$, its variance is

$$\Sigma(k^2 m) - \frac{\Sigma^2(k\alpha m)}{\Sigma(m\alpha^2)}, \qquad (17.104)$$

and a second constraint reduces the variance by a term similar to the second in this expression. The result may be seen from geometrical considerations. We may write

$$\Sigma(kn) = \Sigma\left(k\sqrt{m}\,.\frac{n}{\sqrt{m}}\right) \text{ and}$$

$$\Sigma(\alpha n) = \Sigma\left(\alpha\sqrt{m}\,.\frac{n}{\sqrt{m}}\right),$$

where the variables $\dfrac{n}{\sqrt{m}}$ have unit variance and mean $\sqrt{m}$. Consider the different values of the n's, say s in number, as the co-ordinates in a Euclidean space. The density function of the variables is then symmetrical about a point $(\sqrt{m_1}, \sqrt{m_2}, \ldots \sqrt{m_s})$ to which we transfer the origin. The variance of the unconstrained variables is then equal to the reciprocal of the distance from the origin to the hyperplane $\Sigma(k\sqrt{m}x) = 1$, namely, to $\Sigma(k^2 m)$. But when the constraint is imposed, the variance becomes proportional to the reciprocal of the distance from the origin to the hyperplane in the direction parallel to $\Sigma(\alpha\sqrt{m}x) = 0$ and is hence reduced by the amount

$$\cos^2\phi\, \Sigma(k^2 m),$$

where ϕ is the angle between the planes. This quantity is

$$\frac{\Sigma^2(k\sqrt{m}\,.\,\alpha\sqrt{m})}{\Sigma(k^2 m)\,\Sigma(\alpha^2 m)}\,\Sigma(k^2 m),$$

which gives us the second term in (17.104).

Now for the first linear constraint $\Sigma(n) =$ constant $= n$ we have $\alpha = 1$, and the reducing term is (since $\Sigma(m) = n$ also) :

$$-\frac{1}{n}\Sigma^2(km).$$

For the second constraint we have $\alpha = \dfrac{m'}{m}$ and hence the term is

$$-\frac{\Sigma^2(km')}{\Sigma\left(\dfrac{m'^2}{m}\right)}.$$

Thus the variance of $\Sigma(kn)$ is

$$\Sigma(k^2 m) - \frac{1}{n}\Sigma^2(km) - \frac{\Sigma^2(km')}{\Sigma\left(\dfrac{m'^2}{m}\right)}. \qquad (17.105)$$

Now taking

$$k = \frac{m''}{m} - \frac{m'^2}{m^2}$$

and remembering that

$$\frac{1}{\text{var } t} = \Sigma\left(\frac{m'^2}{m}\right),$$

we see from (17.102) that the loss of information is, for large samples,

$$\frac{\Sigma\left\{\dfrac{1}{m}\left(m'' - \dfrac{m'^2}{m}\right)^2\right\}}{\Sigma\left(\dfrac{m'^2}{m}\right)} - \frac{1}{n}\Sigma\left(\frac{m'^2}{m}\right) - \frac{\Sigma^2\left\{\dfrac{m'}{m}\left(m'' - \dfrac{m'^2}{m}\right)\right\}}{\Sigma^2\left(\dfrac{m'^2}{m}\right)}. \qquad (17.106)$$

By considering the width of the groups as tending to zero we may apply this result also to continuous distributions.

Example 17.16

In the distribution

$$dF = \frac{1}{\pi}\frac{dx}{1 + (x - \theta)^2}, \qquad -\infty \leqslant x \leqslant \infty$$

there is no sufficient estimator, as we have seen. Let us consider the loss of information consequent upon using the maximum likelihood estimator.

We may write for our "expected" value m

$$m = \frac{n}{\pi}\frac{dx}{1 + (x - \theta)^2}.$$

Hence

$$\Sigma\left(\frac{m'^2}{m}\right) = \frac{n}{\pi}\int_{-\infty}^{\infty}\frac{4p^2\,dp}{(1 + p^2)^3} = \frac{n}{2}$$

$$\Sigma\left\{\frac{1}{m}\left(m'' - \frac{m'^2}{m}\right)^2\right\} = \frac{n}{\pi}\int_{-\infty}^{\infty}\frac{4(p^2 - 1)^2\,dp}{(1 + p^2)^5} = \frac{7n}{8}$$

$$\Sigma\left\{\frac{m'}{m}\left(m'' - \frac{m'^2}{m}\right)\right\} = 0.$$

Hence, from (17.106), the loss of information is

$$\frac{7}{4} - \frac{1}{2} + 0 = \frac{5}{4}.$$

The intrinsic accuracy of the original distribution is $\frac{1}{2}$, so the loss of information is equivalent to $2\frac{1}{2}$ observations for large samples. For small samples it will presumably be smaller, since it vanishes for samples of one. The loss by use of the maximum likelihood estimator is therefore very slight and becomes of diminishing importance as the size of the sample increases.

Ancillary Estimators

17.40. Where no sufficient estimator exists no single estimator can avoid the loss of information; but we may take an additional function of the variables which, together with the maximum likelihood estimator, will give an accuracy tending to unity in large samples. By taking a third function we can improve the accuracy still further, and so

on. The process is analogous to approximating to the value of a function (the likelihood function) by ascertaining its differential coefficients at some particular point of the range.

In fact, suppose that, in addition to the estimator which gives $\frac{\partial \log L}{\partial \theta}$ for some value of θ such as t, we also find $\frac{\partial^2 \log L}{\partial \theta^2}$ for that value. The variance of $\frac{\partial \log L}{\partial \theta}$ over values in the neighbourhood of those for which these two are constant is then, to the first approximation, the variance of

$$\tfrac{1}{2}(t-\theta)^2 \frac{\partial^3 \log L}{\partial \theta^3},$$

which has ordinarily a mean value and variance of lower order in n. In particular, if t is the maximum likelihood estimator, so that $\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta=t} = 0$, the value of $\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta=t}$ may provide supplementary information which enables us to approximate more closely to the likelihood function and hence salvage some of the lost information. Such a quantity is accordingly called an *ancillary* estimator. Cf. **17.29** above.

Multivariate Distributions with One Parameter

17.41. We now proceed to consider the extension of some of the foregoing results in two directions: (*a*) where there is more than one variate but still only one parameter, and (*b*) where there is more than one parameter to be estimated.

The former raises no new point of difficulty. To take the bivariate case as an example, if the frequency function is $f(x, y, \theta)$, the likelihood is

$$L = f(x_1, y_1, \theta) \;.\;.\;.\; f(x_n, y_n, \theta) \qquad (17.107)$$

and our maximum likelihood estimator is obtained by maximising L in the usual way.

Example 17.17

To estimate the parameter ρ in samples of n from

$$dF = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\} dx\, dy.$$

We find

$$\log L = \text{constant} - \frac{n}{2}\log(1-\rho^2) - \frac{1}{2(1-\rho^2)}\{\Sigma(x^2) - 2\rho\,\Sigma(xy) + \Sigma(y^2)\},$$

whence, for $\frac{\partial \log L}{\partial \rho} = 0$ we have

$$\frac{n\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}\{\Sigma(x^2) - 2\rho\,\Sigma(xy) + \Sigma(y^2)\} + \frac{1}{1-\rho^2}\Sigma(xy) = 0;$$

reducing to the cubic in ρ,

$$n + \frac{1+\rho^2}{\rho(1-\rho^2)}\Sigma(xy) - \frac{1}{1-\rho^2}\{\Sigma(x^2) + \Sigma(y^2)\} = 0.$$

It is interesting to note that this does *not* yield the product-moment of the sample.

We have, after a little reduction,

$$\frac{\partial^2 \log f}{\partial \rho^2} = \frac{1+\rho^2}{(1-\rho^2)^2} - (x^2 - 2\rho xy + y^2)\frac{1+3\rho^2}{(1-\rho^2)^3} + \frac{4\rho}{(1-\rho^2)^2}\, xy.$$

Since $E(x^2) = E(y^2) = 1$ and $E(xy) = \rho$, we have, for the estimator $\hat{\rho}$,

$$-\frac{1}{n \operatorname{var} \hat{\rho}} = \frac{1+\rho^2}{(1-\rho^2)^2} - \frac{2(1+3\rho^2)}{(1-\rho^2)^2} + \frac{4\rho^2}{(1-\rho^2)^2},$$

whence
$$\operatorname{var} \hat{\rho} = \frac{(1-\rho^2)^2}{n(1+\rho^2)}.$$

This is less (and may be considerably less) than the variance of the sample product-moment in large samples, $(1-\rho^2)^2/n$. The efficiency of the latter is $1/(1+\rho^2)$.

Simultaneous Estimation of Several Parameters

17.42. We now turn to the case when the unknown parameters are more than one in number. To simplify the exposition we shall consider the case of two parameters θ_1 and θ_2, but examples not infrequently arise where more than two have to be estimated—for instance, in the fitting of certain Pearson curves there are four. To fix the ideas, consider the normal distribution

$$dF = \frac{1}{\theta_2 \sqrt{(2\pi)}} \exp\left\{-\frac{1}{2\theta_2^2}(x-\theta_1)^2\right\} dx, \qquad -\infty \leqslant x \leqslant \infty.$$

The likelihood function, except for constants, is given by

$$\log L = -n \log \theta_2 - \frac{1}{2\theta_2^2} \Sigma (x-\theta_1)^2. \quad . \quad . \quad . \quad (17.108)$$

It is natural to generalise our principle of estimation by looking for estimators which shall maximise L for independent simultaneous variations of θ_1 and θ_2, i.e. to require that

$$\frac{\partial \log L}{\partial \theta_1} = 0, \quad \frac{\partial \log L}{\partial \theta_2} = 0. \quad . \quad . \quad . \quad . \quad . \quad (17.109)$$

In our case this leads to

$$\Sigma (x - \theta_1) = 0$$

$$-\frac{n}{\theta_2} + \frac{1}{\theta_2^3} \Sigma (x-\theta_1)^2 = 0,$$

whence for the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$,

$$\hat{\theta}_1 = \frac{1}{n} \Sigma (x) = \bar{x} \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.110)$$

$$\hat{\theta}_2^2 = \frac{1}{n} \Sigma (x-\bar{x})^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.111)$$

Thus the sample mean and variance are estimates of the population mean and variance. We note incidentally that the estimator $\hat{\theta}_2$ is biassed.

17.43. There is one possible source of confusion here which should be removed. If we know θ_1, then $\hat{\theta}_2$ is given by

$$\hat{\theta}_2 = \frac{1}{n} \Sigma (x-\theta_1)^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.112)$$

which is not the same as (17.111), the sample-mean $\bar{x}$ having been replaced by the known

quantity θ_1. Suppose then we estimate θ_1 by $\bar{x}$, as we may do whether we know θ_2 or not, since (17.110) does not contain θ_2. We may then ask, what is the estimator of θ_2 which maximises the likelihood for all samples giving the ascertained value of θ_1, namely, $\bar{x}$?

This is an entirely different question from the one which gave rise to (17.111) and we must not be surprised if it has a different answer. The variations of L from sample to sample are now considered in a certain sub-population for which $\bar{x}$ has a fixed value.

In our particular case the problem can be solved explicitly. The likelihood function can be thrown into the form, with variables $\bar{x}$ and s—

$$L\,d\bar{x}\,ds = \frac{1}{\theta_2}\sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n}{2\theta_2^2}(\bar{x}-\theta_1)^2\right\}$$
$$\times \frac{n^{\frac{1}{2}(n-1)}}{2^{\frac{1}{2}(n-3)}\,\Gamma\{\frac{1}{2}(n-1)\}}\left(\frac{s}{\theta_2}\right)^{n-2}\frac{1}{\theta_2}\exp\left(-\frac{ns^2}{2\theta_2^2}\right)d\bar{x}\,ds, \quad . \quad . \quad (17.113)$$

where s^2 is the sample variance.

If we maximise the likelihood in this form for simultaneous variations of θ_1 and θ_2 we arrive back at (17.110) and (17.111), as of course we must. But if $\bar{x}$ has a fixed value, the distribution of s becomes of one lower degree of freedom. The likelihood is then proportional to the second factor in (17.113), viz.

$$\frac{s^{n-2}}{\theta_2^{n-1}}\exp\left(-\frac{ns^2}{2\theta_2^2}\right),$$

and for variations of θ_2 this is maximised by

$$\hat{\theta}_2^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2. \quad . \quad . \quad . \quad (17.114)$$

This, it may be noticed, is an unbiassed estimator.

17.44. The difference between (17.111) and (17.114) is apt to be confusing, for both are, in a sense, maximum likelihood estimators. The distinction arises from the fact that we are considering the variation of L in two different populations, the first over all samples of size n, the second over the more restricted samples subject to the further constraint $\Sigma(x) = \text{constant}$. The difference when n is large, of course, is quite unimportant, but as a theoretical matter the point has some interest.

Which of the two is employed for practical estimation is a matter of choice. At first sight it may strike the reader as objectionable to use (17.114), because $\bar{x}$ is not known before the sample is drawn, and there are obvious dangers in basing an inference on properties of the sample which are determined *a posteriori*. This objection, however, does not lie in the present case. We make up our mind beforehand that, whatever $\bar{x}$ may turn out to be, we will make an inference in relation to the sub-population of samples determined by it. There is, in fact, no posterior determination of the rule of inference.

17.45. Possibly without realising it, the reader is already accustomed to make an inference of this kind in relation to a sample number. We do not usually determine beforehand what size the sample must be; our results (apart from the distinction between small and large samples, which is another matter) are true for any n, whatever n may turn out to be in practice. In the same way the estimator (17.114) is a maximum likelihood estimator, whatever $\bar{x}$ may turn out to be, $\bar{x}$ being a property of the sample, just as n is.

The fact remains, of course, that (17.111) and (17.114) give different results. Which

is the better? The answer depends on what we require of the estimator. If we wish to choose θ_1 and θ_2 so as to maximise their joint likelihood we choose (17.111). If we wish to select them so that the likelihood is maximised for θ_1 and then, for the observed $\bar{x}$, is maximised for θ_2, we choose (17.114).

17.46. It may be shown that, as for the case of one parameter, the likelihood estimators of several parameters are consistent under very general conditions and tend for large n to be distributed in the multivariate normal form. We omit the proof of these results, which the reader will probably be willing to accept, and proceed to a generalisation of the theorem of **17.26.** Thus :—

(a) If the frequency function $f(x, \theta_1, \theta_2, \ldots \theta_p)$ is continuous in x, and

(b) if in a certain interval containing the true values $\theta_{10}, \theta_{20}, \ldots \theta_{p0}$, $\dfrac{\partial f}{\partial \theta_j}$ is continuous in θ_j for every x, $x^2 \dfrac{\partial f}{\partial \theta_j}$ approaches a continuous function of θ_j for large n, and $\dfrac{\partial f}{\partial \theta_j}$ does not vanish in some interval, then

$$n \operatorname{cov}(\hat{\theta}_j, \hat{\theta}_k) = \frac{\Delta_{jk}}{\Delta}, \qquad (17.115)$$

where Δ is the (Hessian) determinant

$$\Delta = \left| \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta_j}\right)_{\theta_{j0}} \left(\frac{\partial \log f}{\partial \theta_k}\right)_{\theta_{k0}} f\, dx \right| \qquad (17.116)$$

and Δ_{jk} is the minor of the jth row and kth column. When $p = 1$ this reduces to the case of a single parameter.

As n tends to infinity the joint distribution of the maximum likelihood estimators tends to the form

$$f = k \exp\left\{-\frac{n}{2}\Sigma\, g_{jk}\, (\hat{\theta}_j - \theta_{j0})(\hat{\theta}_k - \theta_{k0})\right\}. \qquad (17.117)$$

The theorem will be established if we show that

$$g_{jk} = \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta_j}\right)_{\theta_{j0}} \left(\frac{\partial \log f}{\partial \theta_k}\right)_{\theta_{k0}} dx, \qquad (17.118)$$

for then the values of the variances and covariances of the $\hat{\theta}$'s are as stated in (17.116). (Compare **15.12.**)

Make the transformation

$$q_h = \sum_j A_{hj}\, (\hat{\theta}_j - \theta_{j0}) \qquad (17.119)$$

and choose the A's so that the exponential of (17.117) becomes

$$-\frac{n}{2}\sum_1^p q_h^2.$$

Then
$$g_{jk} = \sum_h A_{hj} A_{hk}. \qquad (17.120)$$

The q's are independent normal variates with variance $1/n$. Hence, from the theorem for the case of a single parameter, already proved, we have

$$\int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial q_h}\right)^2 f\, dx = 1. \qquad (17.121)$$

Further, we have

$$\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial q_h}\frac{\partial \log f}{\partial q_l}\right) f\,dx = 0, \qquad h \neq l, \quad . \quad . \quad . \quad . \quad (17.122)$$

for if we put

$$q_h = \frac{1}{\sqrt{2}}(u_h - u_l)$$

and

$$q_l = \frac{1}{\sqrt{2}}(u_h + u_l)$$

the expression becomes one half of

$$\int_{-\infty}^{\infty} f\,dx \left\{\left(\frac{\partial \log f}{\partial u_h}\right)^2 - \left(\frac{\partial \log f}{\partial u_l}\right)^2\right\},$$

which vanishes since the u's have the same variance as the q's.

Now

$$\left(\frac{\partial \log f}{\partial \theta_j}\right)_{\theta_{j0}} = \Sigma \frac{\partial \log f}{\partial q_h}\left(\frac{\partial q_h}{\partial \theta_j}\right)_{\theta_{j0}} = -\sum_h A_{hj}\frac{\partial \log f}{\partial q_h}$$

Hence

$$\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta_j}\right)_{\theta_{j0}}\left(\frac{\partial \log f}{\partial \theta_k}\right)_{\theta_{k0}} f\,dx = \int_{-\infty}^{\infty}\left(\sum_{h,\,l} A_{hj}A_{lk}\frac{\partial \log f}{\partial q_h}\frac{\partial \log f}{\partial q_l}\right) f\,dx$$

$$= \sum_h A_{hj}A_{hk},$$

in virtue of (17.121) and (17.122),

$$= g_{jk}$$

from (17.120). The theorem follows.

Example 17.18

Let us estimate the five parameters of the bivariate normal form

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\alpha}{\sigma_1}\right)^2 - \frac{2\rho(x-\alpha)(y-\beta)}{\sigma_1\sigma_2} + \left(\frac{y-\beta}{\sigma_2}\right)^2\right\}\right]dx\,dy, \qquad -\infty \leqslant x,\ y \leqslant \infty.$$

It will be found that the partial differential coefficients of $\log L$ yield, on solution, the estimators

$$\hat{\alpha} = \bar{x}, \qquad \hat{\beta} = \bar{y}$$

$$\hat{\sigma}_1^2 = \frac{1}{n}\Sigma(x-\bar{x})^2$$

$$\hat{\rho}\,\hat{\sigma}_1\hat{\sigma}_2 = \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$$

$$\hat{\sigma}_2^2 = \frac{1}{n}\Sigma(y-\bar{y})^2$$

so that for simultaneous estimation the sample means, variances and covariances are estimates of the corresponding parameters.

To evaluate the sampling variances and covariances we have to evaluate integrals of the type

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta_j}\frac{\partial \log f}{\partial \theta_k}\right)dF.$$

These are easily obtainable, being merely functions of moments of different orders.

Example 17.19

Consider the Type III distribution

$$dF = \frac{1}{\sigma\Gamma(\rho)}\left(\frac{x-\alpha}{\sigma}\right)^{\rho-1} \exp\left\{-\left(\frac{x-\alpha}{\sigma}\right)\right\} dx, \qquad \alpha \leqslant x \leqslant \infty.$$

For the likelihood we have

$$\log L = -n\log\sigma - n\log\Gamma(\rho) + (\rho-1)\Sigma\log\left(\frac{x-\alpha}{\sigma}\right) - \Sigma\left(\frac{x-\alpha}{\sigma}\right).$$

The three partial differential coefficients give

$$\begin{aligned} &-(\rho-1)\Sigma\frac{1}{(x-\alpha)} && +\frac{n}{\sigma} = 0 \\ &-\frac{n}{\sigma}\rho && +\frac{1}{\sigma^2}\Sigma(x-\alpha) = 0 \\ &-n\frac{d}{d\rho}\log\Gamma(\rho) && +\Sigma\log\left(\frac{x-\alpha}{\sigma}\right) = 0. \end{aligned}$$

For the Hessian, taking the parameters in the order α, σ, ρ, we have

$$\begin{vmatrix} \dfrac{1}{\sigma^2(\rho-2)} & \dfrac{1}{\sigma^2} & \dfrac{1}{\sigma(\rho-1)} \\ \dfrac{1}{\sigma^2} & \dfrac{\rho}{\sigma^2} & \dfrac{1}{\sigma} \\ \dfrac{1}{\sigma(\rho-1)} & \dfrac{1}{\sigma} & \dfrac{d^2\log\Gamma(\rho)}{d\rho^2} \end{vmatrix}$$

$$= \frac{1}{(\rho-2)\sigma^4}\left\{2\frac{d^2\log\Gamma(\rho)}{d\rho^2} - \frac{2}{\rho-1} + \frac{1}{(\rho-1)^2}\right\} = \Delta.$$

From this the sampling variances are found to be

$$\operatorname{var}\hat{\alpha} = \frac{1}{n\Delta\sigma^2}\left\{\rho\frac{d^2\log\Gamma(\rho)}{d\rho^2} - 1\right\}$$

$$\operatorname{var}\hat{\sigma} = \frac{1}{n\Delta\sigma^2}\left\{\frac{1}{\rho-2}\frac{d^2\log\Gamma(\rho)}{d\rho^2} - \frac{1}{(\rho-1)^2}\right\}$$

$$\operatorname{var}\hat{\rho} = \frac{2}{n\Delta(\rho-2)\sigma^4}.$$

Sufficient Estimators for Several Parameters

17.47. As a natural generalisation from the case of one parameter we shall say that $t_1 \ldots t_p$ are *jointly* sufficient for $\theta_1 \ldots \theta_p$ if, and only if, the likelihood function can be expressed as

$$L(x_1 \ldots x_n, \theta_1 \ldots \theta_p) = L_1(t_1 \ldots t_p, \theta_1 \ldots \theta_p)\, L_2(x_1 \ldots x_n) \qquad (17.123)$$

It evidently does not follow that if $\theta_2 \ldots \theta_p$ are known t_1 is sufficient for θ_1. This will be true only if the function L_1 may itself be factorised, e.g.—

$$L_1(t_1 \ldots t_p, \theta_1 \ldots \theta_p) = L_{11}(t_1, \theta_1 \ldots \theta_p)\, L_{12}(t_2 \ldots t_p, \theta_2 \ldots \theta_p). \qquad (17.124)$$

If a case occurred in which

$$L_1 = L_{11}(t_1, \theta_1)\, L_{12}(t_2, \theta_2) \ldots L_{1p}(t_p, \theta_p) \qquad (17.125)$$

we might say that each t was sufficient for the corresponding θ or that the set of t's was *completely sufficient* for the θ's. Such cases, however, are very rare.

Example 17.20

From (17.113) it is evident that $\bar{x}$ and s are jointly sufficient for m and σ. If σ is known $\bar{x}$ is sufficient for m, but if m is known s is not sufficient for σ. The two are not completely sufficient.

17.48. The properties of sufficient estimators may be proved true, with certain modifications, for several parameters, but we shall not take the subject further except to quote one result.

If $f(x, \theta_1 \ldots \theta_p)$ is continuous and not zero over some continuous range of the θ's, and $\frac{\partial f}{\partial x}$ exists, then it is necessary and sufficient for the existence of a set of jointly sufficient estimators that

$$f = \exp\left\{\sum_{k=1}^{p} A_k X_k + B + Y\right\}, \qquad (17.126)$$

where A_k and B are arbitrary functions of the θ's and X_k and Y of x. (See Koopman, 1936.)

Example 17.21

The Type III distribution of Example 17.19 gives us

$$\log f = -\rho \log \sigma - \log \Gamma(\rho) + (\rho - 1)\log(x - \alpha) - \frac{x - \alpha}{\sigma}.$$

If α is regarded as known, this may be put in the form

$$-\frac{x-\alpha}{\sigma} + (\rho - 1)\log(x - \alpha) - \rho\log\sigma - \log\Gamma(\rho),$$

which is of type (17.126) with

$$\begin{aligned} A_1 &= -\frac{1}{\sigma}, & X_1 &= x - \alpha \\ A_2 &= \rho - 1, & X_2 &= \log(x - \alpha) \\ B &= -\rho\log\sigma - \log\Gamma(\rho). \end{aligned}$$

Thus if α is known, there are sufficient estimators for σ and ρ jointly. It will be clear on inspection that if α is unknown there are no sufficient estimators, even if σ and ρ are known.

Parameters of Location and Scale

17.49. Consider a frequency function expressed in the form

$$dF = g\left(\frac{x-\alpha}{\beta}\right) d\left(\frac{x-\alpha}{\beta}\right) \qquad (17.127)$$

The parameter α may be regarded as locating the distribution and β as determining its scale. In particular the normal distribution may be put in this form. We may write

$$dF = \exp\phi(\xi)\, d\xi = \exp\phi(\xi)\frac{dx}{\beta}, \qquad (17.128)$$

where

$$\xi = \frac{x-\alpha}{\beta} \quad \text{and} \quad \phi(\xi) = \log g(\xi).$$

In samples of n we have

$$\log L = \Sigma\, \phi - n \log \beta,$$

giving for the maximum likelihood estimators

$$\frac{\partial \log L}{\partial \alpha} = -\frac{1}{\beta} \Sigma\, \phi' = 0 \quad . \quad . \quad . \quad . \quad . \quad (17.129)$$

$$\frac{\partial \log L}{\partial \beta} = -\frac{1}{\beta} (\Sigma\, \phi'\, \xi + n) = 0, \quad . \quad . \quad . \quad . \quad (17.130)$$

whence we may solve for $\hat{\alpha}$ and $\hat{\beta}$.

For the variances and covariance we find

$$E\left(\frac{\partial^2 \log f}{\partial \alpha^2}\right) = E\left(\frac{\phi''}{\beta^2}\right) = -E\left(\frac{\partial \log f}{\partial \alpha}\right)^2$$

$$E\left(\frac{\partial^2 \log f}{\partial \beta^2}\right) = E\left\{\frac{1}{\beta^2} (\phi''\, \xi^2 + 2\, \phi'\, \xi + 1)\right\}$$

$$= E\left\{\frac{1}{\beta^2} (\phi''\, \xi^2 - 1)\right\} = -E\left(\frac{\partial \log f}{\partial \beta}\right)^2$$

$$E\left(\frac{\partial^2 \log f}{\partial \alpha\, \partial \beta}\right) = E\left\{\frac{1}{\beta^2} (\phi' + \phi''\, \xi)\right\}$$

$$= E\left\{\frac{1}{\beta^2} \phi''\, \xi\right\} = -E\left(\frac{\partial \log f}{\partial \alpha}\, \frac{\partial \log f}{\partial \beta}\right),$$

and the Hessian of (17.116) becomes

$$\begin{vmatrix} -E\left(\dfrac{\phi''}{\beta^2}\right) & -E\left(\dfrac{\phi''\, \xi}{\beta^2}\right) \\ -E\left(\dfrac{\phi''\, \xi}{\beta^2}\right) & -E\left(\dfrac{\phi''\, \xi^2 - 1}{\beta^2}\right) \end{vmatrix} \quad . \quad . \quad . \quad (17.131)$$

from which the variances and covariance of $\hat{\alpha}$ and $\hat{\beta}$ may be determined in the usual way.

In (17.131) it would be a great convenience if the quantity $-E\left(\dfrac{\phi''\, \xi}{\beta^2}\right)$ vanished, for then $\hat{\alpha}$ and $\hat{\beta}$ would be independent. By a suitable choice of origin we can, in fact, ensure that this is so. Put

$$\zeta = \xi - \frac{E\,(\phi''\, \xi)}{E\,(\phi'')}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.132)$$

Then

$$E\,(\phi''\, \xi) = E\left\{(\phi''\, \zeta) + \phi'' \frac{E\,(\phi''\, \xi)}{E\,(\phi'')}\right\}$$

$$= E\,(\zeta \phi'' + \xi \phi''),$$

so that

$$E\,(\phi''\, \zeta) = 0.$$

With this origin we have for the variances of the (uncorrelated) variables $\hat{\alpha}$ and $\hat{\beta}$,

$$\operatorname{var} \hat{\alpha} = -\frac{\beta^2}{nE\,(\phi'')} \quad . \quad . \quad . \quad . \quad . \quad . \quad (17.133)$$

$$\operatorname{var} \hat{\beta} = -\frac{\beta^2}{n\,\{E\,(\phi''\, \zeta^2) - 1\}}. \quad . \quad . \quad . \quad . \quad (17.134)$$

The point of location so defined, namely, as that for which $\hat{\alpha}$ and $\hat{\beta}$ are uncorrelated, has been called by Fisher the *centre of location.*

Example 17.22

For the normal distribution

$$dF = \frac{1}{\beta\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2\right\} dx$$

we have

$$\phi = -\tfrac{1}{2}\xi^2$$

$$E(\phi'') = -1 \qquad \text{and} \qquad E(\phi''\xi) = 0.$$

Hence $\zeta = \xi$, and the origin chosen is itself the centre of location. From (17.133) and (17.134) we find the familiar results (for large samples)

$$\text{var}\,\hat{\alpha} = \text{var}\,\bar{x} = \frac{\beta^2}{n}$$

$$\text{var}\,\hat{\beta} = \text{var}\,s = \frac{\beta^2}{2n},$$

with $\bar{x}$ and s uncorrelated.

Example 17.23

Consider again the Type III distribution

$$dF = \frac{1}{\Gamma(\rho)}\left(\frac{x-\alpha}{\beta}\right)^{\rho-1} \exp\left\{-\frac{x-\alpha}{\beta}\right\} d\left(\frac{x-\alpha}{\beta}\right), \qquad \alpha \leqslant x \leqslant \infty, \quad \rho > 2$$

where we assume ρ known. The condition $\rho > 1$ is required to ensure the vanishing of the frequency function at the extremity $x = \alpha$, and $\rho > 2$ to ensure the convergence of some of the mean values.

Here

$$\phi = \text{constant} - \xi + (\rho - 1)\log\xi.$$

Hence

$$E(\phi'') = E\left(-\frac{\rho-1}{\xi^2}\right) = -\frac{1}{\rho-2}$$

$$E(\xi\phi'') = E\left(-\frac{\rho-1}{\xi}\right) = -1$$

$$E(\xi^2\phi'') = E(-\rho+1) = -(\rho-1).$$

Thus

$$\zeta = \xi - (\rho - 2).$$

The centre of location is distant $(\rho - 2)$ to the right of the start of the distribution. In terms of ζ we have

$$\phi = \text{constant} - \zeta - (\rho-2) + (\rho-1)\log(\zeta+\rho-2)$$

$$\phi' = -1 + \frac{\rho-1}{\zeta+\rho-2} \qquad \phi'' = -\frac{(\rho-1)}{(\zeta+\rho-2)^2}$$

$$E(\phi'') = -1/(\rho-2)$$

$$E(\phi''\zeta^2 - 1) = -2.$$

Hence

$$\text{var}\,\hat{\alpha} = \frac{\beta^2(\rho-2)}{n}$$

$$\text{var}\,\hat{\beta} = \frac{\beta^2}{2n}.$$

Efficiency of the Method of Moments

17.50. In previous chapters we have fitted distributions of the Pearson type to other distributions by identifying lower moments. We were there mainly concerned with the properties of populations only and no question of the reliability of estimates arose. If, however, we regard the data as a *sample* from a population, the question arises whether fitting by moments provides the most efficient estimators of the unknown parameters. As we shall see presently, in general it does not.

Consider a parent form dependent on four parameters. If the maximum likelihood estimators of these parameters are to be obtained in terms of linear functions of the moments (as in the fitting of Pearson curves), we must have

$$\frac{\partial \log L}{\partial \theta} = a_0 + a_1 \Sigma(x) + a_2 \Sigma(x^2) + a_3 \Sigma(x^3) + a_4 \Sigma(x^4) \quad . \quad (17.135)$$

and consequently

$$f(x, \theta_1, \theta_2, \theta_3, \theta_4) = \exp\{b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4\}, \quad . \quad (17.136)$$

where the b's depend on the θ's. This is the most general form for which the method of moments gives maximum likelihood estimators. The b's are, of course, conditioned by the fact that the total frequency shall be unity and the distribution function converge.

Without loss of generality we may take $b_1 = 0$. If, then, the other b's vanish except b_0 and b_2 the distribution is normal and the method of moments is most-efficient. In other cases, (17.136) does not yield a Pearson distribution except as an approximation. For example,

$$\frac{d \log f}{dx} = 2b_2 x + 3b_3 x^2 + 4b_4 x^3.$$

If b_3 and b_4 are small this is approximately

$$\frac{d \log f}{dx} = \frac{2b_2 x}{1 - \dfrac{3b_3}{2b_2} x - \dfrac{2b_4}{b_2} x^2}, \quad . \quad . \quad . \quad . \quad (17.137)$$

which is one form of the equation defining Pearson distributions (cf. **6.2**). Only when b_3 and b_4 are small compared with b_2 can we expect the method of moments to give estimates of high efficiency.

17.51. A detailed discussion of the efficiency of moments in determining the parameters of a Pearson distribution has been given by Fisher (1921*a*). We will here quote only one of the results by way of illustration.

We found in Example 17.19 that the variance for large samples of the maximum likelihood estimator $\hat{\rho}$ is given by

$$\operatorname{var} \hat{\rho} = \frac{2}{n\left\{2\dfrac{d^2 \log \Gamma(\rho)}{d\rho^2} - \dfrac{2}{\rho - 1} + \dfrac{1}{(\rho - 1)^2}\right\}}$$

or, if $p = \rho - 1$, by

$$\operatorname{var} \hat{\rho} = \frac{2}{n\left\{2\dfrac{d^2 \log \Gamma(1 + p)}{dp^2} - \dfrac{2}{p} + \dfrac{1}{p^2}\right\}}. \quad . \quad . \quad (17.138)$$

Now for large p,*

$$\frac{d^2}{dp^2}\log \Gamma(1+p) = \frac{d^2}{dp^2}\left\{\frac{1}{2}\log 2\pi + (p+\tfrac{1}{2})\log p - p + \frac{1}{12p} - \frac{1}{360p^3} + \frac{1}{1260p^5} - \dots\right\}$$

We then find

$$2\frac{d^2}{dp^2}\log \Gamma(1+p) - \frac{2}{p} + \frac{1}{p^2} = \frac{1}{3}\left\{\frac{1}{p^3} - \frac{1}{5p^5} + \frac{1}{7p^7} - \dots\right\}$$

and hence, approximately,

$$\text{var}\,\hat{p} = \frac{6}{n}(p^3 + \tfrac{1}{5}p). \qquad (17.139)$$

If we estimate the parameters by equating sample-moments to the appropriate moments in terms of parameters, we find

$$\begin{aligned} \alpha + \sigma\rho &= m_1 \\ \sigma^2\rho &= m_2 \\ 2\rho\sigma^3 &= m_3 \end{aligned}$$

so that, whatever α and σ may be,

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{4}{\rho}, \qquad (17.140)$$

where b_1 is the sample value of β_1. Now for estimation by the method of moments (cf. **9.22**),

$$\text{var}\,b_1 = \frac{\beta_1}{n}(4\beta_4 - 24\beta_2 + 36 + 9\beta_1\beta_2 - 12\beta_3 + 35\beta_1),$$

which for the present distribution is readily seen to reduce to

$$\text{var}\,b_1 = \frac{\beta_1^2}{n}.\frac{6(\rho+1)(\rho+5)}{\rho}. \qquad (17.141)$$

Hence, from (17.140) we have for ρ, estimated by the method of moments,

$$\begin{aligned} \text{var}\,\rho &= \frac{\rho^4}{16}\text{var}\,b_1 \\ &= \frac{6}{n}\rho(\rho+1)(\rho+5). \end{aligned}$$

For large ρ the efficiency of this estimator is then, from (17.139) with $\rho = 1 + p$,

$$E = \frac{p^3 + \frac{1}{5}p}{(p+1)(p+2)(p+6)},$$

which is evidently short of unity in many cases. When p exceeds 38·1 ($\beta_1 = 0{\cdot}102$) the efficiency is over 80 per cent. For $p = 19$ ($\beta_1 = 0{\cdot}20$) it is 65 per cent. For $p = 4$ a more exact calculation based on the tables of the trigamma function $\dfrac{d^2\log\Gamma(1+p)}{dp^2}$ shows that the efficiency is only 22 per cent.

* The series for the log Γ function is given in most books on advanced calculus, e.g. J. Edwards, *Integral Calculus*, vol. 2, article 942.

NOTES AND REFERENCES

The greater part of this chapter is based on the researches of R. A. Fisher, the main papers being those of 1921*a*, 1925*b* and 1934*a*. The idea of maximising likelihood may be traced back to Gauss and was considered by Edgeworth, but may be regarded as beginning to exercise an influence on statistical theory only with the publication of Fisher's first paper in 1912.

The theorem giving the limiting variances and covariances of maximum likelihood estimates was proved (incorrectly) by Karl Pearson and Filon in 1898 before it was realised that it applied only to maximum likelihood. The necessary correction was given by Edgeworth (1908) and Fisher (1921*a*), but rigorous proofs were not available until the work of Hotelling (1930) and Doob (1934*a* and *b*, 1935, 1936). In the text we have followed Hotelling's treatment.

The inefficiency of moments in fitting distributions, pointed out by Fisher (1921*a*), has led to some controversy, for which see Koshal (1933, 1935), Myers (1934), Elderton and Hansmann (1934), K. Pearson (1936), and Fisher (1937*a*). The reader who pursues this subject so far as to read any one of these papers should read them all.

For work on sufficient estimators see Koopman (1936) and Pitman (1936, 1937*b*), who independently obtained the general form of distribution admitting such estimators. The theorem that sufficient estimators have the property **17.17** is due to Fisher, rigorous proofs being provided by Neyman (1935*a*) and Dugué (1936*a*). Reference should also be made to papers by Bartlett (1936*a*, *b*, 1937*c*, 1938*b*, 1939*a*, 1940) on the problem of several parameters and what he calls " conditional " statistics, i.e. those similar to s^2 when $\bar{x}$ or some other function of the sample values is regarded as known. See also Neyman and Pearson (1936*a*).

Among recent papers, that by Pitman (1939*a*) on parameters of scale and location, and that by Welch (1939*c*) on the distribution of maximum likelihood estimates, are noteworthy.

Geary (1942) has recently proved a remarkable generalisation of the theorem that in large samples maximum-likelihood estimators have minimum variance in the case of one parameter. In fact, for several parameters the maximum likelihood estimators minimise the " generalised variance " as defined in Chapter 28.

EXERCISES

17.1. If t is a most-efficient estimator and t' a less-efficient estimator with efficiency E, and if the correlation of t and t' is ρ, show by considering the estimator t'' defined by

$$(1 + E - 2\rho\sqrt{E})\, t'' = (1 - \rho\sqrt{E})\, t + (E - \rho\sqrt{E})\, t'$$

that $\rho = \sqrt{E}$ (for in the contrary case var t'' > var t).

(Fisher, 1925*b*.)

17.2. If in n trials of an event with probability p there are x successes, show that a maximum likelihood estimator of p is x/n. Find its sampling variance and show that it is sufficient.

17.3. Show that the distribution

$$dF = \tfrac{1}{2} \exp \{-\mid x - \theta \mid\}\, dx, \qquad -\infty \leqslant x \leqslant \infty$$

has a likelihood function for a sample of n which is a maximum at the median if n is odd and between the $(n/2)$th and $(n/2+1)$th members if n is even.

17.4. For the distribution of the previous exercise show that for a sample of $(2m+1)$ members the median has an accuracy

$$\frac{(m+1)(2m+1)}{(m-1)}\left\{1-\frac{(2m)!}{2^{2m-1}(m!)^2}\right\}.$$

Hence, as m tends to infinity, the loss of information tends to $4\sqrt{(m/\pi)}-4$. Thus, although the median is most-efficient the loss of information in large samples does not tend to a constant.

(Fisher, 1925*b*.)

17.5. Show that if a most-efficient estimator A and a less-efficient estimator B tend to joint normality for large samples, $B-A$ tends to zero correlation with A.

Show that the error in B may be regarded as composed (for large samples) of two parts which are independent, the error in A and the error in $B-A$. (The first may be regarded as sampling error, necessarily inherent in the problem of estimation, the second as error due to the inefficiency of the estimator.)

(Fisher, 1925*b*.)

17.6. Show that the distribution of the median in a sample of $(2m+1)$ observations from the population

$$dF=\frac{1}{\pi}\frac{dx}{1+(x-\theta)^2}, \qquad -\infty\leqslant x\leqslant\infty$$

is given by

$$dF=\frac{(2m+1)!}{(m!)^2\pi^{2m+1}}\left(\frac{\pi^2}{4}-\phi^2\right)^m\frac{dx}{1+(x-\theta)^2},$$

where $\tan\phi=x-\theta$ and $|\phi|\leqslant\frac{1}{2}\pi$.

Show hence that the accuracy of the median is

$$\frac{(2m+1)!}{(m!)^2\pi^{2m+1}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left\{2m\phi\cos^2\phi+\left(\frac{\pi^2}{4}-\phi^2\right)\sin 2\phi\right\}^2\left(\frac{\pi^2}{4}-\phi^2\right)^{m-2}d\phi$$

$$=\tfrac{1}{2}+\frac{3m(2m+1)}{2(m-1)\pi^2}+\frac{(m+\frac{1}{2})!}{2m-1}\left(\frac{2}{\pi}\right)^{m+\frac{1}{2}}\left\{\frac{2m}{m-1}J_{m-\frac{3}{2}}(\pi)-2J_{m+\frac{1}{2}}(\pi)\right\}$$

$$+\frac{(m+\frac{1}{2})!}{2m-1}\left(\frac{1}{\pi}\right)^{m+\frac{1}{2}}\left\{\frac{2m}{m-1}J_{m-\frac{3}{2}}(2\pi)-\frac{2m+3}{2}J_{m+\frac{1}{2}}(2\pi)\right\}$$

where $J_n(z)$ is the Bessel function of order n and in particular $J_{\frac{1}{2}}(\pi)=J_{\frac{1}{2}}(2\pi)=0$, $J_{\frac{3}{2}}(\pi)=\frac{\sqrt{2}}{\pi}$, $J_{\frac{3}{2}}(2\pi)=-\frac{1}{\pi}$, and

$$J_{n+1}=\frac{2n}{z}J_n-J_{n-1}$$

(Fisher, 1925*b*.)

17.7. Show that the most general continuous distribution for which the maximum likelihood estimator of a parameter θ is the geometric mean of the sample is

$$f(x, \theta) = \left(\frac{x}{\theta}\right)^{\theta \frac{\partial \psi}{\partial \theta}} \exp \{\psi(\theta) + \zeta(x)\},$$

where ψ is an arbitrary function of θ, and ζ of x. Show further that the corresponding distribution giving the harmonic mean is

$$f(x, \theta) = \exp\left[\frac{1}{x}\left\{\theta \frac{\partial \psi}{\partial \theta} - \psi\right\} - \frac{\partial \psi}{\partial \theta} + \zeta(x)\right]$$

(Keynes, *J.R.S.S.* (1911), **74**, 323.)

17.8. Show that, if m is known, the estimator

$$s = \left\{\frac{1}{n} \Sigma (x - m)^2\right\}^{\frac{1}{2}}$$

is sufficient for σ in samples of n from

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2\sigma^2}(x - m)^2\right\} dx,$$

and find its distribution by the method of **17.31**.

17.9. By considering the distribution

$$dF = e^{-(x-\theta)}\, dx, \qquad \theta \leqslant x \leqslant \infty$$

show that the three forms of (17.97) are not necessarily equivalent when the range contains the parameter to be estimated.

(Pitman, 1936.)

17.10. Show that if the frequency function is continuous and is zero at an extreme which is a function of θ, there still exists a maximum to the intrinsic accuracy, defined as $E\left(\frac{\partial \log f}{\partial \theta}\right)^2$.

(Pitman, 1936.)

17.11. By considering the distribution

$$dF = \frac{2x}{2\theta + 1}, \qquad \theta \leqslant x \leqslant \theta + 1$$

show that the intrinsic accuracy is $4n^2/(2\theta + 1)^2$. Show further that the largest member of the sample is sufficient for θ and that its distribution is

$$dF = \alpha(x)\, dx = \frac{2nx\,(x^2 - \theta^2)^{n-1}}{(2\theta + 1)^n}\, dx.$$

Hence show that

$$E\left(\frac{\partial \log \alpha}{\partial \theta}\right)^2 = \frac{4n^2(\theta + 1)^2}{(2\theta + 1)^2} + \frac{4n\theta^2}{(n-2)(2\theta + 1)^2},$$

so that the mean value in this case is greater than the intrinsic accuracy.

(Pitman, 1936.)

17.12. If the frequency function of an estimator t is Φ its accuracy is $E\left\{\frac{1}{\Phi}\left(\frac{\partial\Phi}{\partial\theta}\right)^2\right\}$. If every possible sample with frequency ϕ gave a different value of t the accuracy would be $E\left\{\frac{1}{\phi}\left(\frac{\partial\phi}{\partial\theta}\right)^2\right\}$ and would be independent of t. Show that the difference in accuracy may be expressed as

$$E\left\{\phi\left(\frac{1}{\phi}\frac{\partial\phi}{\partial\theta}-\frac{1}{\Phi}\frac{\partial\Phi}{\partial\theta}\right)^2\right\}$$

and hence is not negative.

Hence show that the efficiency as defined in **17.36** cannot exceed unity, at least if the range is independent of θ.

(Fisher, 1925*b*.)

17.13. Show that

$$dF=\frac{1}{\pi}\frac{\theta_2\,dx}{\theta_2^2+(x-\theta_1)^2},\qquad -\infty\leqslant x\leqslant\infty$$

does not admit of a sufficient estimator for either parameter if the other is known, or a pair of jointly sufficient estimators if both are unknown.

(Koopman, 1936.)

17.14. Show that if a distribution admits a sufficient estimator for either of two parameters when the other is known, it admits of a pair of jointly sufficient estimators when both parameters are unknown.

(Koopman, 1936.)

17.15. Show that the centre of location of the Type IV distribution

$$dF\propto e^{-\nu\tan^{-1}(x-\alpha)/\beta}\left\{1+\left(\frac{x-\alpha}{\beta}\right)^2\right\}^{-\frac{\rho+2}{2}}dx,\qquad -\infty\leqslant x\leqslant\infty$$

where ν and ρ are assumed known, is distant $\frac{\nu\beta}{\rho+4}$ to the left of the mode of the distribution.

(Fisher, 1921*a*.)

17.16. For the distribution

$$dF=\frac{dx}{\theta_2},\qquad \theta_1-\frac{\theta_2}{2}\leqslant x\leqslant\theta_1+\frac{\theta_2}{2}$$

show that, in large samples, the mean tends to the form

$$dF=\frac{1}{\theta_2}\sqrt{\frac{6n}{\pi}}\exp\left(-\frac{6n\bar{x}^2}{\theta_2^2}\right)d\bar{x}.$$

Show further that the distribution of the centre of the sample, say c (the mean of the two extreme values), tends to

$$dF=\frac{n}{\theta_2}\exp\left\{-\frac{2n}{\theta_2}|\,c\,|\right\}dc.$$

Hence
$$\frac{\text{var}\,c}{\text{var}\,\bar{x}}=\frac{6}{n},$$

so that the centre is a far better estimator of location than the mean for this distribution.

(Fisher, 1921*a*.)

17.17. Show that for the Type I distribution

$$dF = \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}\, dx, \qquad 0 \leqslant x \leqslant 1$$

the geometric mean of the sample values x and that of the values $(1-x)$ are jointly sufficient for the estimation of p and q.

17.18. Show that all the Pearson distributions have sufficient estimators for some of the parameters if the others are assumed known, and ascertain which are the parameters concerned for each type.

17.19. For the distribution of Exercise 17.15 show that the intrinsic accuracy for α is

$$\frac{1}{\beta^2} \frac{(\rho+1)(\rho+2)(\rho+4)}{(\rho+4)^2+\nu^2},$$

and that the efficiency of the method of moments in locating the curve is

$$\frac{\rho^2(\rho-1)\{(\rho+4)^2+\nu^2\}}{(\rho+1)(\rho+2)(\rho+4)(\rho^2+\nu^2)}.$$

(Fisher, 1921*a*.)

CHAPTER 18

ESTIMATION : MISCELLANEOUS METHODS

Minimum Variance

18.1. We have seen in the previous chapter that under certain general conditions the maximum likelihood estimator is most-efficient for large samples, and that for finite samples it leads to sufficient estimators where such exist. Sufficient estimators themselves contain all the information in the sample about the parameter under estimate. What we have not shown, however, is that maximum likelihood estimators have minimum variance in finite samples.

We now consider the subject from a slightly different standpoint. Instead of beginning with the criteria of efficiency and sufficiency and showing that they lead to certain minimal properties, we shall examine the class of estimators which (*a*) are unbiassed and (*b*) have minimum variance. The minimal property is here taken as the starting-point.

18.2. Consider, then, a frequency function $f(x, \theta)$, and as usual let us write $L = f(x_1, \theta) \ldots f(x_n, \theta)$. Then, writing $\int dx$ for the n-fold integral over the range of the x's, we have to find $t = t(x_1, \ldots x_n)$ such that

$$\int_{-\infty}^{\infty} t\, L\, dx = \theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.1)$$

$$\int_{-\infty}^{\infty} (t-\theta)^2 L\, dx = \text{minimum}. \quad . \quad . \quad . \quad . \quad (18.2)$$

The first equation may also be written

$$\int_{-\infty}^{\infty} (t-\theta) L\, dx = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.3)$$

The problem of finding t is one of the familiar problems in the Calculus of Variations. The minimal value of (18.2) has to be found subject to the condition (18.1), which is equivalent to

$$\int_{-\infty}^{\infty} t \frac{\partial L}{\partial \theta} dx = 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.4)$$

provided that the range of f is independent of θ or that f vanishes at any extreme which depends on θ.

If 2λ is an unspecified parameter (which may depend on θ but not on the x's) the problem is equivalent to finding an unconditioned minimum of

$$\int_{-\infty}^{\infty} \left\{ (t-\theta)^2 L - 2\lambda t \frac{\partial L}{\partial \theta} \right\} dx. \quad . \quad . \quad . \quad . \quad . \quad (18.5)$$

The solution is *

$$\frac{\partial}{\partial t} \left\{ (t-\theta)^2 L - 2\lambda t \frac{\partial L}{\partial \theta} \right\} = 0$$

* See, for example, J. Edwards, *Integral Calculus*, vol. 2, article 1504, or A. R. Forsyth, *Calculus of Variations*, article 15. Since the expression to be minimised does not contain $\frac{\partial t}{\partial x}$, the Euler equation for a stationary value to the integral $\int V\, dx$ reduces to $\frac{\partial V}{\partial t} = 0$. The derivation of (18.7) is not,

or

$$(t - \theta) L - \lambda \frac{\partial L}{\partial \theta} = 0. \quad . \quad . \quad . \quad . \quad . \quad (18.6)$$

We then have

$$t = \theta + \frac{\lambda}{L} \frac{\partial L}{\partial \theta}$$

$$= \theta + \lambda \frac{\partial \log L}{\partial \theta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.7)$$

where t is a function of the x's but not of θ. Thus there exists a t satisfying our conditions if we can express $\frac{\partial \log L}{\partial \theta}$ in the form

$$\frac{\partial \log L}{\partial \theta} = \frac{t - \theta}{\lambda}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.8)$$

This is a necessary and sufficient condition, except that it gives only *stationary* values of (18.2) which might, for instance, be maxima instead of minima. This is not a point, however, which need detain us from the statistical viewpoint, troublesome as it is to the mathematician.

Example 18.1

To estimate θ in the normal population

$$dF = \frac{1}{\sigma \sqrt{(2\pi)}} \exp \left\{ - \frac{1}{2\sigma^2} (x - \theta)^2 \right\} dx, \qquad -\infty \leqslant x \leqslant \infty$$

where σ is assumed known.

We have

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\sigma^2} (\bar{x} - \theta).$$

This can be put in the form (18.8) by taking

$$\bar{x} = t \qquad \text{and} \qquad \lambda = \frac{\sigma^2}{n},$$

and hence $\bar{x}$ is the required estimator. We note that it has minimum variance for any *n in the class of unbiassed estimators of* θ.

Example 18.2

To estimate θ in

$$dF = \frac{1}{\pi} \frac{dx}{1 + (x - \theta)^2}, \qquad -\infty \leqslant x \leqslant \infty.$$

We have

$$\frac{\partial \log L}{\partial \theta} = 2 \Sigma \left\{ \frac{x - \theta}{1 + (x - \theta)^2} \right\}.$$

This cannot be put in the form (18.8) and the method fails. There is no estimator which is unbiassed and has minimum variance.

however, without its difficulties, and I think some conditions have been accidentally suppressed in the Aitken-Silverstone method. I understand that Dr. Leon Solomon, working with Dr. Aitken, has obtained a proof which depends on the fact that L shall be the product of n independent frequency functions. But for the war the point would doubtless have been cleared up by now, but at present it remains open.

18.3. Integrating (18.8) with respect to θ we have

$$\log L = \alpha(\theta)(t-\theta) + \beta(\theta) + \sum_j \gamma(x_j),$$

where α, β, γ are arbitrary functions (apart from the fact that the two former depend on λ). Hence

$$\begin{aligned}\log f(x, \theta) &= A(\theta)(t-\theta) + B(\theta) + C(x) \\ &= p(\theta)\, t(x) + q(\theta) + r(x), \text{ say.} \end{aligned} \qquad (18.9)$$

Comparing this with (17.83), we see that the method of minimum variance will give a solution only if there exists a sufficient estimator. This explains the success of the method in Example 18.1 (where $\bar{x}$ is sufficient) and its failure in Example 18.2 (where no sufficient estimator exists).

18.4. In the method of maximum likelihood it makes no difference to the final result whether we estimate for a parameter θ or for some other parameter χ functionally related to θ. For

$$\frac{\partial \log L}{\partial \theta} = \frac{\partial \log L}{\partial \chi}\frac{\partial \chi}{\partial \theta}$$

and the two sides of the equation vanish together. In the method of minimum variance, however, there is an interesting difference.

Suppose we wish to estimate θ in

$$dF = \frac{1}{\sqrt{(2\pi\theta)}} \exp\left(-\frac{1}{2}\frac{x^2}{\theta}\right) dx, \qquad -\infty \leqslant x \leqslant \infty.$$

We have

$$\frac{\partial \log L}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2}\frac{\Sigma(x^2)}{\theta^2},$$

and this may be put in the form (18.8) with

$$t = \frac{1}{n}\Sigma(x^2) \quad \text{and} \quad \lambda = \frac{2\theta^2}{n}.$$

If, however, we consider the parallel problem of estimating σ in

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx, \qquad -\infty \leqslant x \leqslant \infty$$

we find

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\Sigma(x^2)}{\sigma^3},$$

which cannot be put in the form (18.8). We thus reach the peculiar result that the method will provide an estimator for σ^2 but not for σ. It follows that in general we may have to estimate, not θ itself, but some function of θ, say $\tau(\theta)$.

18.5. If a minimum-variance estimator exists for some $\tau(\theta)$ we must have

$$\frac{\partial \log L}{\partial \tau} = \frac{t-\tau}{\lambda(\tau)},$$

which is equivalent to

$$\frac{\partial \log L}{\partial \theta} = \frac{\frac{\partial \tau}{\partial \theta}(t-\tau)}{\lambda(\theta)}. \qquad (18.10)$$

We estimate t by putting it equal to τ and thus we shall have, for the estimator,

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{t=\tau} = 0. \qquad \qquad (18.11)$$

This is equivalent to the equation of maximum likelihood. The two are not, however, identical. Maximum likelihood is not concerned with the existence of the function λ. Minimum variance takes the function as fundamental, and when it exists the solution (which is the same as the maximum likelihood solution) has minimum variance for all n in the class of unbiassed estimators, not merely for large n.

18.6. Let us suppose that θ is the parameter (transformed if necessary) for which the estimating function is θ itself. Then we have for the minimum-variance estimator t

$$\text{var } t = \int_{-\infty}^{\infty} (t - \theta)^2 L \, dx,$$

which, on substitution from (18.8), yields

$$\text{var } t = \int_{-\infty}^{\infty} \lambda^2 \left(\frac{\partial \log L}{\partial \theta}\right)^2 L \, dx \qquad \qquad (18.12)$$

$$= -\lambda^2 \int_{-\infty}^{\infty} \left(\frac{\partial^2 \log L}{\partial \theta^2}\right) L \, dx, \qquad \qquad (18.13)$$

if the range is independent of θ or f vanishes at any extreme dependent on θ.

Now from (18.8) we find

$$\frac{\partial^2 \log L}{\partial \theta^2} = (t - \theta) \frac{\partial}{\partial \theta}\left(\frac{1}{\lambda}\right) - \frac{1}{\lambda},$$

and hence, substituting in (18.13) and remembering that $\int_{-\infty}^{\infty} (t - \theta) L \, dx = 0$, we find

$$\text{var } t = -\lambda^2 \int_{-\infty}^{\infty} \left(-\frac{1}{\lambda}\right) L \, dx$$

$$= \lambda. \qquad \qquad (18.14)$$

The variance of the minimum-variance estimator is thus simply the parameter λ. It also follows from (18.13) that

$$\frac{1}{\text{var } t} = -\int_{-\infty}^{\infty} \left(\frac{\partial^2 \log L}{\partial \theta^2}\right) L \, dx$$

$$= -n E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right), \qquad \qquad (18.15)$$

so that the result we reached in Chapter 17, as a limiting form for large n, is now seen to be exact for finite n under present conditions.

Example 18.3

To estimate θ in the Type III form

$$dF = \frac{1}{\Gamma(\rho)\,\theta^\rho} x^{\rho-1} e^{-x/\theta} \, dx, \qquad 0 \leqslant x \leqslant \infty, \quad \rho > 1,$$

where ρ is assumed known.

We have

$$\frac{\partial \log L}{\partial \theta} = -\frac{n\rho}{\theta} + \frac{n\bar{x}}{\theta^2}$$

which is of the form (18.8) if

$$t = \frac{\bar{x}}{\rho} \quad \text{and} \quad \lambda = \frac{\theta^2}{n\rho}.$$

Thus t is the minimum-variance estimator and has variance $\frac{\theta^2}{n\rho}$ for finite n, even though the distribution is not normal. (Compare Example 17.8.)

18.7. We may readily determine what function $\tau(\theta)$ should be taken as the estimating function. Taking the general form from (18.9),

$$\log f(x, \theta) = p(\theta)\, t(x) + q(\theta) + r(x),$$

we have

$$\log L = p\, \Sigma\, t(x) + nq + \Sigma\, r(x)$$

$$\frac{\partial \log L}{\partial \tau} = \frac{\partial p}{\partial \tau} \Sigma(t) + n\frac{\partial q}{\partial \tau}$$

$$= n\frac{\partial p}{\partial \tau}\left(\frac{1}{n}\Sigma(t) + \frac{\partial q}{\partial p}\right). \qquad (18.16)$$

Hence, if

$$\tau = -\frac{\partial q}{\partial p} = -\frac{\partial q}{\partial \theta}\Big/\frac{\partial p}{\partial \theta} \qquad (18.17)$$

we have

$$\frac{\partial \log L}{\partial \tau} = \frac{\frac{1}{n}\Sigma(t) - \tau}{1/n\, \frac{\partial p}{\partial \tau}}, \qquad (18.18)$$

which is of the required form provided that

$$\frac{1}{\lambda} = n\frac{\partial p}{\partial \tau}. \qquad (18.19)$$

Example 18.4

Consider again the estimation of σ in

$$dF = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Here

$$\log f = -\tfrac{1}{2}\log(2\pi) - \log\sigma - \tfrac{1}{2}\frac{x^2}{\sigma^2},$$

whence

$$p(\sigma) = -\frac{1}{2\sigma^2}, \quad t(x) = x^2, \quad q = -\log\sigma.$$

Thus the appropriate value of τ, from (18.17), is

$$\tau = -\frac{\partial q}{\partial \sigma}\Big/\frac{\partial p}{\partial \sigma}$$

$$= \sigma^2,$$

which is thus determined as our estimating function. For the variance of the estimator of τ we have

$$\lambda = 1/n \frac{\partial p}{\partial \tau} = \frac{2\sigma^4}{n},$$

the estimator itself being $\frac{1}{n} \Sigma (x^2)$.

Minimum χ^2

18.8. We now turn to consider another principle which has been suggested for providing estimators. If the data are grouped into cells with expected frequency typified by λ_j and observed frequency by l_j, then the function

$$\chi^2 = \Sigma \frac{(l_j - \lambda_j)^2}{\lambda_j} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18.20)$$

$$= \Sigma \left(\frac{l_j^2}{\lambda_j} \right) - n,$$

where

$$n = \Sigma (\lambda_j) = \Sigma (l_j) \quad . \quad . \quad . \quad . \quad . \quad (18.21)$$

can, as we saw in Chapter 12, be used as a measure of closeness of fit. The method of minimum χ^2 adopts this standpoint (which is, of course, arbitrary in the logical sense) and attempts to determine the parameters λ such that χ^2 is a minimum.

In practice the method is not very easy to apply because of the difficulty of expressing the λ's in terms of the parameter under estimate, θ. For some illustrations reference may be made to Kirstine Smith (1916). We shall not consider the method at length here for two reasons :—

(*a*) it may be shown that for large samples the minimum-χ^2 estimator tends to the maximum-likelihood estimator ;

(*b*) there is a modification of the method, considered below, which is much easier to apply.

18.9. For samples of fixed size n the distribution of the quantities l_j is multinomial, and we have for the likelihood function

$$L = \frac{n\,!}{\prod\limits_j (l_j\,!)} \prod_j \left(\frac{\lambda_j}{n} \right)^{l_j}$$

$$= \frac{n\,!}{\prod\limits_j (l_j\,!)} \Pi \left(\frac{l_j}{n} \right)^{l_j} \Pi \left(\frac{\lambda_j}{l_j} \right)^{l_j}. \quad . \quad . \quad . \quad . \quad . \quad (18.22)$$

Thus

$$\log L = \text{constant} + \Sigma\, l_j \log \left(\frac{\lambda_j}{l_j} \right). \quad . \quad . \quad . \quad . \quad . \quad (18.23)$$

Now for large samples we may put

$$\lambda_j = l_j + a_j\, n^{\frac{1}{2}},$$

where a_j is finite and therefore small compared with l_j; $| a_j n^{\frac{1}{2}} | < l_j$; and $\Sigma (a_j) = 0$.

Hence, from (18.23),

$$\log L = k + \Sigma\, l_j \log\left(1 + \frac{a_j\, n^{\frac{1}{2}}}{l_j}\right)$$
$$= k - \tfrac{1}{2}\Sigma \frac{n\, a_j^2}{l_j} + O\,(n^{-\frac{1}{2}})$$
$$= k - \tfrac{1}{2}\Sigma \frac{(\lambda_j - l_j)^2}{l_j} + O\,(n^{-\frac{1}{2}}). \qquad (18.24)$$

Now write

$$\chi'^2 = \Sigma \frac{(\lambda_j - l_j)^2}{l_j}$$
$$= \Sigma \frac{\lambda_j^2}{l_j} - n. \qquad (18.25)$$

Then we see that, to order $n^{-\frac{1}{2}}$, L is maximised by minimising χ'^2. This latter quantity is not the same as χ^2 because the denominator terms are l's instead of λ's. However, for large n the difference is of order $n^{-\frac{1}{2}}$, for

$$\chi^2 - \chi'^2 = \Sigma\, (\lambda_j - l_j)^2 \left\{\frac{1}{\lambda_j} - \frac{1}{l_j}\right\}$$
$$= \Sigma \frac{(\lambda_j - l_j)^2}{l_j}\left\{\left(1 + \frac{a_j\, n^{\frac{1}{2}}}{l_j}\right)^{-1} - 1\right\}$$
$$= -\,\Sigma \frac{(\lambda_j - l_j)^2}{l_j^2}\, a_j\, n^{\frac{1}{2}} + \ldots$$
$$= O\,(n^{-\frac{1}{2}}).$$

Hence, to order $n^{-\frac{1}{2}}$ the estimates obtained by minimising either χ^2 or χ'^2 will be equivalent to maximising L.

18.10. The advantage of using χ'^2 instead of χ^2 in practice resides in the fact that the denominators in the former are integral. However, if there are any empty cells (i.e. those for which $l_j = 0$) the formula (18.25) requires some modification.

In the likelihood function, if $l_j = 0$, $\left(\frac{\lambda_j}{n}\right)^{l_j} = 1$ for all λ_j. The substitution

$$\lambda_j = l_j + a_j\, n^{\frac{1}{2}}$$

will give us, for the empty cells, a term in (18.24) equal to $-\,\Sigma\, a_j\, n^{\frac{1}{2}} = -\,\Sigma\, \lambda_j = M$, say. Hence we have

$$\chi'^2 = \Sigma \frac{(\lambda_j - l_j)^2}{l_j} + 2M, \qquad (18.26)$$

where the summation takes place over occupied cells and M is the sum of the theoretical frequencies λ in the empty cells.

Example 18.5

As an example (Jeffreys, 1941) we consider a case where the maximum likelihood estimator is known, so that a comparison may be made with the result given by minimum χ'^2.

Col. (2) of the following table shows the frequency of women in the first class of Part II

of the Mathematical Tripos from 1910 to 1938 inclusive. Assuming that this distribution follows the Poisson distribution $\frac{e^{-\theta}\theta^j}{j!}$, to estimate θ.

(1) Number of firsts, j	(2) Frequency l_j	(3) λ_j			(4) χ'^2		
		$\theta = 1$	$\theta = 1{\cdot}5$	$\theta = 2$	$\theta = 1$	$\theta = 1{\cdot}5$	$\theta = 2$
0	6	10·7	6·5	3·9	3·7	0·0	0·7
1	8	10·7	9·7	7·9	0·9	0·4	0·0
2	11	5·3	7·3	7·9	3·0	1·2	0·9
3	3	1·8	3·6	5·2	0·5	0·1	1·6
4	0	0·5	1·4	2·6	—	—	—
5	1	0·1	0·4	1·0	0·8	0·4	0·0
over 5	0	0·0	0·1	0·5	$2M = 1{\cdot}0$	$2M = 3{\cdot}0$	$2M = 6{\cdot}2$
TOTALS	29				9·9	5·1	9·4

The sample mean (a sufficient estimator of θ) is in this case $44/29 = 1{\cdot}52$ with a standard error $\sqrt{\frac{\bar{x}}{n}} = 0{\cdot}23$.

To apply minimum χ'^2 we have to express the theoretical frequencies in terms of θ. This results in an unmanageable equation if we then substitute in χ'^2. Instead we calculate the minimum by finding χ'^2 for some trial values of θ (in this case 1, 1·5 and 2) and then interpolating.

The expectations λ for the three selected values of θ are shown in column (3) of the table and the corresponding χ'^2 in column (4). It is found that, writing $\theta = 1{\cdot}5 + \phi$, the values of χ'^2 may be represented by the quadratic

$$\chi'^2 = 5{\cdot}1 - 0{\cdot}5\phi + 18{\cdot}2\phi^2.$$

The minimum of this is given by $\phi = 0{\cdot}01$, and hence our estimate of θ is 1·51, very close to the value of 1·52 given by the maximum likelihood estimator.

18.11. On theoretical grounds there seems no reason to use minimum χ^2 instead of maximum likelihood. The method has some practical value, however, where the maximum likelihood equations are difficult to solve. We can usually follow the device of the example just given, find χ^2 or χ'^2 for some trial values of the parameter, and approximate to the value which minimises χ^2 or χ'^2. Whether this is easier than finding the maximum likelihood estimate in the same sort of way depends on the circumstances of the case, but it may well be so when the frequency function is a tabulated integral, so that expected frequencies for specified parameter-values can be readily obtained.

18.12. In the manner of **17.39** we can estimate the loss of information occasioned by the use of minimum χ^2. We have, for the minimum of χ^2,

$$\frac{\partial}{\partial\theta}\Sigma\frac{(l-\lambda)^2}{\lambda} = 0,$$

which reduces to

$$\Sigma \frac{l^2 - \lambda^2}{\lambda^2} \frac{\partial \lambda}{\partial \theta} = 0. \qquad (18.27)$$

Since $\frac{l + \lambda}{\lambda}$ tends to the constant value 2 for large samples, this is equivalent to the maximum likelihood equation

$$\Sigma \frac{l - \lambda}{\lambda} \frac{\partial \lambda}{\partial \theta} = 0, \qquad (18.28)$$

confirming that maximum likelihood and minimum χ^2 give the same results in the limit.

Since

$$l^2 - \lambda^2 = 2\lambda (l - \lambda) + (l - \lambda)^2$$

the deviation of $\frac{\partial \log L}{\partial \theta}$ from its mean is

$$\tfrac{1}{2} \Sigma \frac{l^2 - \lambda^2}{\lambda^2} \frac{\partial \lambda}{\partial \theta} - \tfrac{1}{2} \Sigma \frac{(l - \lambda)^2}{\lambda^2} \frac{\partial \lambda}{\partial \theta}, \qquad (18.29)$$

the first term vanishing on summation. As in **17.39** we find the variance of this quantity within samples for which $\frac{\partial \log L}{\partial \theta}$ is constant. We have

$$\text{var}\, \Sigma\, k (l - \lambda)^2 = 2\, \Sigma (k^2 \lambda^2) - \frac{2}{n} \Sigma^2 (k\lambda^2) - 2 \frac{\Sigma^2 (k\lambda'^2)}{\Sigma^2 \left(\frac{\lambda'^2}{\lambda}\right)},$$

and on substituting $k = \frac{1}{2} \frac{\lambda'}{\lambda^2}$ we find

$$\tfrac{1}{2} \Sigma \left(\frac{\lambda'^2}{\lambda^2}\right) - \tfrac{1}{2} \frac{\Sigma^2 \left(\frac{\lambda'^3}{\lambda^2}\right)}{\Sigma^2 \left(\frac{\lambda'^2}{\lambda}\right)}, \qquad (18.30)$$

giving the loss of information.

As the sample size increases, this quantity remains finite. It is interesting to observe, however, that as the number of classes increases it also increases without limit, indicating that minimum χ^2 breaks down for fine grouping.

"*Inverse*" *Probability*

18.13. According to Bayes' theorem (**7.24**), if $h(\theta)\, d\theta$ is the prior probability of θ, the posterior probability is given by

$$P(\theta \mid x_1, \ldots x_n) = L(x_1, \ldots x_n, \theta)\, h(\theta)\, d\theta\,. \qquad (18.31)$$

It is then easy to determine the "most probable" value of θ by maximising $L\, h(\theta)$ if we know $h(\theta)$. The principles of inference with which we have been concerned up to the present do not require the notion of the probability of θ and, even if they did, would not give any guide to the nature of the function $h(\theta)$. In fact, to an adherent of the frequency theory of probability, the prior probability of θ requires the distribution of θ in some form, and if θ is merely an unknown constant it has no distribution (except the trivial one that $f = 1$ when θ takes its true value and $f = 0$ elsewhere). The alternative school of thought assumes the existence of $h(\theta)$ as denoting a prior measure of belief, but, in order to find

the most probable value of θ, has to make some further assumption as to its values comparable to Bayes' postulate that for a finite range h is a constant.

We have already noted that on this assumption the maximisation of L is equivalent to finding the value of θ with the greatest posterior probability. It is also interesting to note that, whatever the form of $h(\theta)$, maximum likelihood tends to give the same estimator as the method of maximising posterior probability for large n. In fact, for the maximisation of P in (18.31) we have

$$\frac{\partial \log P}{\partial \theta} = \frac{\partial \log L}{\partial \theta} + \frac{\partial \log h}{\partial \theta} = 0. \quad . \quad . \quad . \quad . \quad (18.32)$$

In ordinary cases the variance of $\frac{\partial \log L}{\partial \theta}$ is of order n, whereas the second term is independent of n. In the limit, therefore, the second term is negligible and we are reduced to the likelihood equation

$$\frac{\partial \log L}{\partial \theta} = 0.$$

Least Squares

18.14. The method of least squares bears an analogy to minimum χ^2. Suppose we have an expression depending on a number of unknown parameters $\theta_1 \ldots \theta_p$ and certain observed values x. This can be thrown into a form such as

$$k(x, \theta_1 \ldots \theta_p) = 0, \quad . \quad . \quad . \quad . \quad (18.33)$$

where k is a given function (not a frequency function). If we have n values of x and $n > p$ it is not possible to solve the n resulting equations of type (18.33) for the θ's. We then consider the "residuals" $k(x_j, \theta_1 \ldots \theta_p)$, and the principle of least squares states that the values of $\theta_1 \ldots \theta_p$ are to be chosen so that

$$\Sigma \{k(x_j, \theta_1 \ldots \theta_p)\}^2 = \text{minimum}, \quad . \quad . \quad . \quad (18.34)$$

or, in other words, so as to satisfy the p equations

$$\sum_j \frac{\partial}{\partial \theta_l} \{k^2(x_j, \theta_1 \ldots \theta_p)\} = 0, \qquad l = 1 \ldots p. \quad . \quad . \quad (18.35)$$

18.15. Consider the case when the residuals are all distributed normally with variance σ^2. The logarithm of the likelihood is then (except for constants)—

$$\log L = -n \log \sigma - \frac{1}{2\sigma^2} \Sigma k^2(x_j, \theta_1 \ldots \theta_p) \quad . \quad . \quad . \quad (18.36)$$

and this is clearly maximised by minimising the sum (18.34). In this case, then, the method of least squares is equivalent to the method of maximum likelihood. In other cases it may give different results, and the justification for using it then becomes more or less empirical.

18.16. The most important case occurring in statistical theory of the use of the method of least squares concerns regression equations. We have already seen that the coefficients of regression are, in effect, determined so as to minimise the sum of squares of residuals (cf. **15.2**). We also know that, for the multiple normal distribution, residuals from the population regression lines are, in fact, normally distributed (**15.13**). For normal

variation, therefore, the method of least squares is equivalent to maximum likelihood so far as concerns the simultaneous estimation of regression coefficients.

18.17. This is a convenient point to prove a theorem (due to Gauss) which in one form or another is constantly occurring in statistical theory, particularly in connection with the normal distribution. Suppose we have a population (not necessarily normal) in which the regression of one variate y on the others x_0 $(=1)$, $x_1 \ldots, x_p$ is given by

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p. \qquad (18.37)$$

The x's may be correlated among themselves and, in the extreme case, functionally related, so that this case includes that of curvilinear regression for our present purposes. Suppose that we have a sample of n values, where $n > p$. Denoting by Σ summation over these n values, we determine the estimates of the β's by minimising the sum of squares, e.g.

$$\Sigma (y - \beta_0 - \beta_1 x_1 - \ldots - \beta_p x_p)^2.$$

Suppose that $b_0 \ldots b_p$ are the solutions of this process. Then our regression formula is

$$y - b_0 - b_1 x_1 - \ldots - b_p x_p = 0. \qquad (18.38)$$

The observed residuals, obtained by substituting the observed values in this equation, are typified by

$$e = y - b_0 - b_1 x_1 \ldots - b_p x_p, \qquad (18.39)$$

whereas the "real" residuals are typified by

$$\varepsilon = y - \beta_0 - \beta_1 x_1 \ldots - \beta_p x_p. \qquad (18.40)$$

We proceed to compare the sampling variances of e and ε and to show that

$$\text{var}\,\varepsilon = \frac{n}{n-p-1}\,\text{var}\,e, \qquad (18.41)$$

provided that the residuals are uncorrelated.

Let us transform the observed values of the x's to new values $\xi_0, \xi_1 \ldots \xi_p$ (n for each) such that

$$\left.\begin{aligned} \Sigma (x_j \xi_k) &= 1 \qquad j = k \\ &= 0 \qquad j \neq k \\ \Sigma (\xi_k y) &= b_k \end{aligned}\right\} \qquad (18.42)$$

This involves, for each ξ, $p + 1$ equations in n unknowns and is therefore possible in general. We then have

$$\begin{aligned} -\Sigma \xi_k (\varepsilon - e) &= \Sigma \xi_k \{(\beta_0 - b_0) + (\beta_1 - b_1) x_1 + \ldots (\beta_p - b_p) x_p\} \\ &= \beta_k - b_k. \end{aligned}$$

But
$$\begin{aligned} \Sigma \xi_k e &= \Sigma (\xi_k y) - \Sigma \xi_k \{b_0 + b_1 x_1 + \ldots b_p x_p\} \\ &= b_k - b_k = 0. \end{aligned}$$

Hence
$$\beta_k - b_k = -\Sigma \xi_k \varepsilon. \qquad (18.43)$$

Now
$$\begin{aligned} -\Sigma e (\varepsilon - e) &= \Sigma \{y - b_0 - \ldots - b_p x_p\} \{(\beta_0 - b_0) + \ldots (\beta_p - b_p) x_p\} \\ &= 0, \end{aligned}$$

since the summations give terms the vanishing of which determines the b's. Hence

$$\begin{aligned} \Sigma \varepsilon^2 - \Sigma e^2 &= \Sigma (\varepsilon - e) \varepsilon \\ &= \underset{j}{S} (b_j - \beta_j) \Sigma x_j \varepsilon, \end{aligned}$$

where S denotes summation over the $(p+1)$ values of j,

$$\begin{aligned} &= S\,\Sigma\,\xi_j\,\varepsilon\,\Sigma\,x_j\,\varepsilon \\ &= S\,\{\Sigma\,\xi_j\,x_j\,\varepsilon^2\} + \text{cross-product terms in } \varepsilon, \\ &= S\,\varepsilon^2 + \text{cross-product terms.} \end{aligned}$$

When we take expectations the cross-product terms vanish since the residuals are uncorrelated. Hence

$$E\,(\Sigma\,\varepsilon^2) - E\,(S\,\varepsilon^2) = E\,\Sigma\,e^2,$$

or

$$(n - p - 1)\,\text{var}\,\varepsilon = n\,\text{var}\,e, \qquad . \quad . \quad . \quad (18.44)$$

from which (18.41) follows at once.

For normal variation we shall consider this result from a slightly different viewpoint in Chapter 22.

NOTES AND REFERENCES

The approach to minimum-variance estimators through the Calculus of Variations is due to Aitken and Silverstone (1942). For minimum χ^2 see K. Smith (1916) and R. A. Fisher (1922*a*, 1925*b*). For the modification χ'^2 see Jeffreys (1938*b*, 1939*b*, 1941).

A method of estimation essentially depending on the median has been proposed for use in quality control, but its value is as yet problematical. For an account of the technique see Simon (1941).

EXERCISES

18.1. From the property that the variance of a minimum-variance estimator is equal to λ show that the most general distribution for which the sample mean is a sufficient estimator is

$$f\,(x,\,\theta) = c\,(x,\,\sigma)\,\exp\left\{-\frac{1}{2\sigma^2}\,(x-\theta)^2\right\},$$

where c is an arbitrary function and σ^2 is the variance of f.

Hence show that no Pearson curve other than the normal admits the sample-mean as a sufficient estimator, but that a Gram-Charlier series may do so.

(Aitken and Silverstone, 1942.)

18.2. If the function λ exists and

$$\alpha\,(\theta) = \int_0^\theta \frac{d\theta}{\lambda\,(\theta)},$$

show that the variance of the estimator t is

$$-\frac{1}{n}\,\frac{\partial^2 q}{\partial \alpha^2},$$

where q is the function of **18.7**. (Aitken and Silverstone, 1942.)

18.3. If a population $(p+q)^4$ is regarded as distributed in 5 classes, show that the intrinsic accuracy is $\dfrac{4n}{pq}$. Show further that the loss of information through estimating p from minimum χ^2 is

$$\frac{5}{p^2\,q^2}\,(3p^2 - 2pq + 3q^2) - \frac{(p-q)^2}{2p^2\,q^2}\,(p^4 - 2p^3\,q + 18p^2\,q^2 - 2pq^3 + q^4)^2.$$

This is least when $p = q$ and is then equivalent to the loss of 5 observations.

(Fisher, 1925*b*.)

CHAPTER 19

CONFIDENCE INTERVALS

19.1. In the previous two chapters we have been concerned with methods which will provide an estimate of the value of one or more unknown parameters ; and the methods gave functions of the sample values—the estimators—which, for any given sample, provided a unique estimate. It was of course fully recognised that the estimate might differ from the parameter in any particular case, and hence that there was a margin of uncertainty. The extent of this uncertainty was expressed in terms of the sampling variance of the estimator. With the somewhat intuitional approach which has served our purpose up to this point, we say that it is probable that θ lies in the range $t \pm \sqrt{\text{var } t}$, very probable that it lies in the range $t \pm 2\sqrt{\text{var } t}$, and so on. In short, what we have done is in effect to locate θ in a range and not at a particular point, although we have regarded one point in the range, viz. t itself, as having a claim to be considered as the " best " estimate of θ.

19.2. In the present chapter we shall examine the logic of this procedure more closely and look at the problem of estimation from a different point of view. We now abandon attempts to estimate θ by a function which, for a specified sample, gives a unique number. Instead we shall consider merely the specification of a range in which θ lies. We shall not attempt to specify whereabouts in the interval the value of θ really is ; all values in the range have an equal claim to be taken as the " true " value. Nor shall we assess the probability that θ lies in the interval in the sense that θ is regarded as a random variable. In fact, in the frequency theory of probability θ is not a random variable (except trivially in that the frequency of θ is unity when it takes the true value and is zero elsewhere). Nevertheless, probability plays an essential part in the determination of the interval and in the degree of confidence we have that it " covers " θ.

Case of one Unknown Parameter

19.3. Consider in the first place a population dependent on a single unknown parameter θ and suppose that we are given a random sample of n values $x_1 \ldots x_n$ from the population. Let z be a statistic dependent on the x's and on θ, whose sampling distribution is independent of θ. (The examples given below will show that in some cases at least such a statistic may be found.) Then, given any probability α, we can find a value z_1 such that

$$\int_{-\infty}^{z_1} dF(z) = \alpha,$$

and this is true whatever the value of θ. In the notation of the theory of probability we shall then have

$$P(z \leqslant z_1 \mid \theta) = \alpha. \qquad (19.1)$$

Now it may happen that the inequality $z \leqslant z_1$ can be transformed to the form $\theta \leqslant t_1$ or $\theta \geqslant t_1$, where t_1 is some function depending on the value z_1 and the x's but not on θ. For instance, if $z = \bar{x} - \theta$ we shall have

$$\bar{x} - \theta \leqslant z_1$$

and hence

$$\theta \geqslant \bar{x} - z_1.$$

If this transformation can be made we then have, from (19.1),

$$P(\theta \leqslant t_1 \mid \theta) = \alpha. \qquad (19.2)$$

More generally, suppose that we can find a function t_1, depending on α and the x's but not on θ, such that (19.2) is true for all θ. Then we may use this equation in probability to make certain statements about θ.

19.4. Note, in the first place, that we cannot assert that the probability is α that θ does not exceed a constant t_1. This statement (in the frequency theory of probability) can only relate to the variation of θ in a population of θ's, and in general we do not know that θ varies at all. If it is merely an unknown constant then the probability that $\theta \leqslant t_1$ is either unity or zero. We do not know which of these values is correct, but we do know that one of them is correct.

We therefore look at the matter in another way. Although θ is not a random variable, t_1 is and will vary from sample to sample. Consequently, if we *assert* that $\theta \leqslant t_1$ in each case presented for decision, we shall be right in a proportion α of the cases in the long run. The statement that the probability of θ is less than or equal to some assigned value has no meaning except in the trivial sense already mentioned; but the statement that a statistic t_1 is greater than or equal to θ (whatever θ happens to be) has a definite probability α of being correct. If therefore we make it a rule to assert the inequality $\theta \leqslant t_1$ for any sample values which arise, we have the assurance of being right in a proportion α of the cases "on the average" or "in the long run."

This idea is basic to the theory of confidence intervals which we proceed to develop, and the reader should satisfy himself that he has grasped it.

19.5. To simplify the exposition we have considered only a single quantity t_1 and the statement that $\theta \leqslant t_1$. In practice, however, we usually seek for two quantities t_0 and t_1, such that

$$P\{t_0 \leqslant \theta \leqslant t_1 \mid \theta\} = \alpha, \qquad (19.3)$$

and make the assertion that θ lies in the range t_0 to t_1. These quantities are known as the Lower and Upper Confidence Limits respectively. They depend only on α and the sample values. For any fixed α the totality of values of t_0 and t_1 for different samples determine a field within which θ is asserted to lie. This field is called the Confidence Belt or Region of Acceptance. We shall give a graphical representation of the idea below. The number α is called the Confidence Coefficient.

Example 19.1

Suppose we have a sample of n from the normal population with unit variance

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp\{-\tfrac{1}{2}(x-\mu)^2\}\,dx, \qquad -\infty \leqslant x \leqslant \infty.$$

The distribution of means $\bar{x}$ will be

$$dF = \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\} d\bar{x}, \qquad -\infty \leqslant \bar{x} \leqslant \infty.$$

From the tables of the normal integral we know that the probability of a positive deviation from the mean not greater than twice the standard deviation is 0·97725. We have then—

$$P\left\{\bar{x} - \mu \leqslant \frac{2}{\sqrt{n}} \,|\mu\right\} = 0{\cdot}97725,$$

which is equivalent to

$$P\left\{\bar{x} - \frac{2}{\sqrt{n}} \leqslant \mu \mid \mu\right\} = 0{\cdot}97725.$$

Thus, if we assert that μ is greater than or equal to $\bar{x} - 2/\sqrt{n}$ we shall be right in about 97·725 per cent. of the cases.

Similarly we have

$$P\left\{\bar{x} - \mu \geqslant -\frac{2}{\sqrt{n}} \mid \mu\right\} = P\left\{\mu \leqslant \bar{x} + \frac{2}{\sqrt{n}} \mid \mu\right\} = 0{\cdot}97725.$$

Hence, combining the two results,

$$P\left\{\bar{x} - \frac{2}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + \frac{2}{\sqrt{n}} \mid \mu\right\} = 2\,(0{\cdot}97725) - 1 = 0{\cdot}9545.$$

Hence, if we assert that μ lies in the range $\bar{x} \pm 2/\sqrt{n}$ we shall be right in about 95·45 per cent. of the cases in the long run.

Conversely, given the confidence coefficient we can easily find from the tables of the normal integral the deviation d such that $P\left\{\bar{x} - \frac{d}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + \frac{d}{\sqrt{n}}\right\} = \alpha$. For instance, if $\alpha = 0{\cdot}8$, $d = 1{\cdot}28$, so that if we assert that μ lies in the range $\bar{x} \pm 1{\cdot}28/\sqrt{n}$ the odds are 4 to 1 that we shall be right.

The reader to whom this approach is new will probably ask : but is this not a roundabout way of using the standard error to set limits to an estimate of the mean ? In a way, it is. In effect, what we have done in this example is to show how the use of the standard error of the mean in normal samples may be justified on logical grounds without appeal to new principles of inference other than those incorporated in the theory of probability itself. In particular we make no use of Bayes' postulate.

Another point of interest in this example is that the upper and lower confidence limits derived above are equidistant from the mean $\bar{x}$. This is not by any means necessary, and it is easy to see that we can derive any number of alternative limits for the same confidence coefficient α. Suppose, for instance, we take $\alpha = 0{\cdot}9545$, and select two numbers α_0 and α_1, which obey the condition

$$(\alpha_0 + \alpha_1 - 1) = 0{\cdot}9545,$$

say $\alpha_0 = 0{\cdot}9645$ and $\alpha_1 = 0{\cdot}99$. From the tables of the normal integral we have

$$P\left\{\bar{x} - \mu \leqslant \frac{2{\cdot}326}{\sqrt{n}} \mid \mu\right\} = 0{\cdot}99$$

$$P\left\{\bar{x} - \mu \geqslant -\frac{1{\cdot}806}{\sqrt{n}} \mid \mu\right\} = 0{\cdot}9645,$$

and hence

$$P\left\{\bar{x} - \frac{2{\cdot}326}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + \frac{1{\cdot}806}{\sqrt{n}} \mid \mu\right\} = 0{\cdot}9545.$$

Thus, with the same confidence coefficient we can assert that μ lies in the range $\bar{x} - 2/\sqrt{n}$ to $\bar{x} + 2/\sqrt{n}$, or in the range $\bar{x} - 2{\cdot}326/\sqrt{n}$ to $\bar{x} + 1{\cdot}806/\sqrt{n}$. In either case we shall be right in about 95·45 per cent. of the cases.

We note that in the first case the range is $4/\sqrt{n}$ units and in the second case it is $4{\cdot}132/\sqrt{n}$ units. Other things being equal, we should choose the first set of limits since

they locate the parameter in a narrower range. We shall consider this point in more detail below. It does not always happen that there is an infinity of possible confidence limits or, if there is, that any simple rule of choice between them can be formulated.

Graphical Representation

19.6. In a number of simple cases, including that of the previous example, the confidence limits can be represented in a useful graphical form. We take two orthogonal axes, OX relating to the observed $\bar{x}$ and OY to μ (see Fig. 19.1).

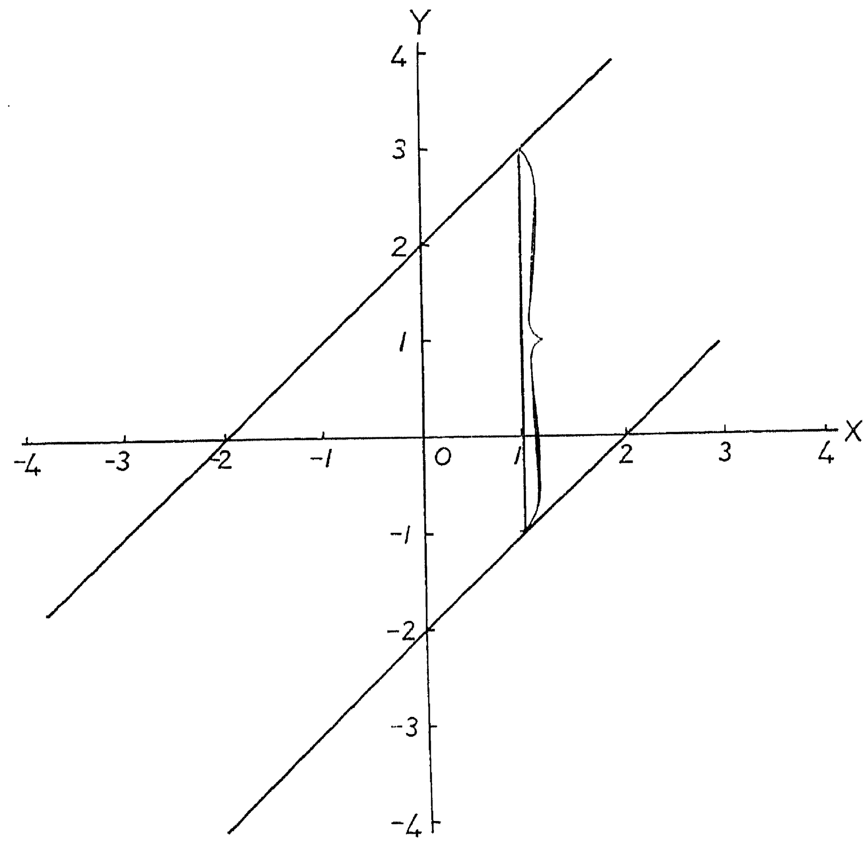


FIG. 19.1.

The two straight lines shown have as their equations

$$\mu = \bar{x} + 2, \qquad \mu = \bar{x} - 2.$$

Consequently, for any point between the lines,

$$\bar{x} - 2 \leqslant \mu \leqslant \bar{x} + 2.$$

Hence, if for any observed $\bar{x}$ we read off the two ordinates on the lines corresponding to that value we obtain the two confidence limits. The vertical interval between the limits is the confidence range (shown in the diagram for $\bar{x} = 1$), and the total zone between the lines is the confidence belt. We may refer to the two lines as the Upper and Lower Confidence lines respectively.

This example relates to the somewhat trivial case $n = 1$. For different values of n there will be different confidence lines, all parallel to $\mu = \bar{x}$. They may be shown on a single diagram for selected values of n, and a figure so constructed provides a useful method of reading off confidence limits in practical work.

Central and Non-central Intervals

19.7. In Example 19.1 the sampling distribution on which the confidence intervals were based was symmetrical, and hence, by taking equal deviations from the mean, we reached equal areas of the frequency function as α_0 and α_1. In general we cannot achieve this result with equal deviations, and subject always to the condition $\alpha_0 + \alpha_1 - 1 = \alpha$ the two quantities may be chosen arbitrarily.

If α_0 and α_1 are taken to be equal, we shall say that the intervals are *central.* In such a case we have

$$P(t_0 \leqslant \theta) = P(\theta \leqslant t_1) = \frac{1+\alpha}{2}. \qquad (19.4)$$

In the contrary case the intervals will be called *non-central.*

19.8. In the absence of other considerations it is usually convenient to employ central intervals, but circumstances sometimes arise in which non-central intervals are more serviceable. Suppose, for instance, we are estimating the proportion of some drug in a medicinal preparation and the drug is toxic in large doses. We must then clearly err on the safe side, an excess of the true value over our estimate being more serious than a deficiency. In such a case we might prefer to take α_1 very near to unity or even equal to unity, so that

$$P(\theta \leqslant t_1) = 1$$
$$P(t_0 \leqslant \theta) = \alpha,$$

and we are *certain* that θ is not greater than t_1.

Again, if we are estimating the proportion of viable seed in a sample of material that is to be placed on the market, we are more concerned with the accuracy of the lower limit than that of the upper limit, for a deficiency of germination is more serious than an excess from the grower's point of view. In such circumstances we should probably take α_0 as large as conveniently possible so as to be nearer to certainty about the minimum value of viability. This kind of situation often arises in the specification of the quality of a manufactured product, the seller wishing to guarantee a minimum standard but being much less concerned with whether his product exceeds expectation.

19.9. On a somewhat similar point, it may be remarked that in certain circumstances it is enough to know that $P\{t_0 \leqslant \theta \leqslant t_1 \mid \theta\}$ *exceeds* some quantity α. We then know that in asserting θ to lie in the range t_0 to t_1 we shall be right in *at least* a proportion α of the cases. Mathematical difficulties in ascertaining confidence limits exactly for given α, or theoretical difficulties when the distribution is discontinuous may, for example, lead us to be content with the inequality rather than the equality of (19.3).

Example 19.2

To find confidence intervals for the parent proportion ϖ of successes in sampling for attributes.

In samples of n the distribution of successes is given by the binomial $(\chi + \varpi)^n$. We will determine the limits for the case $n = 20$ and confidence coefficient 0·95.

We require in the first instance the distribution function of the binomial, which is obtainable from Table 5.2 (vol. I, p. 119). Summing the number of successes and dividing by 10,000, we find from that table the following :—

Proportion of Successes p	$\varpi = 0{\cdot}1$	$\varpi = 0{\cdot}2$	$\varpi = 0{\cdot}3$	$\varpi = 0{\cdot}4$	$\varpi = 0{\cdot}5$
0·00	0·1216	0·0115	0·0008	—	—
0·05	0·3918	0·0691	0·0076	0·0005	—
0·10	0·6770	0·2060	0·0354	0·0036	0·0002
0·15	0·8671	0·4114	0·1070	0·0159	0·0013
0·20	0·9569	0·6296	0·2374	0·0509	0·0059
0·25	0·9888	0·8042	0·4163	0·1255	0·0207
0·30	0·9977	0·9133	0·6079	0·2499	0·0577
0·35	0·9997	0·9678	0·7722	0·4158	0·1316
0·40	1·0001	0·9900	0·8866	0·5955	0·2517
0·45	1·0002	0·9974	0·9520	0·7552	0·4119
0·50	—	0·9994	0·9828	0·8723	0·5881
0·55	—	0·9999	0·9948	0·9433	0·7483
0·60	—	1·0000	0·9987	0·9788	0·8684
0·65	—	—	0·9997	0·9934	0·9423
0·70	—	—	0·9999	0·9983	0·9793
0·75	—	—	—	0·9996	0·9941
0·80	—	—	—	0·9999	0·9987
0·85	—	—	—	—	0·9998
0·90	—	—	—	—	1·0000
0·95	—	—	—	—	—

The final figures may be a unit or two in error owing to rounding up, but that need not bother us to the degree of approximation here considered. Values for $\varpi = 0{\cdot}6$ to $0{\cdot}9$ may be obtained by symmetry.

We note in the first place that the variate p is discontinuous. On the other hand we are prepared to consider any value of ϖ in the range 0 to 1. For given ϖ we cannot in general find limits to p for which α is exactly 0·95 ; but we will take p to be the nearest multiple of 0·05 which gives confidence coefficients at least equal to 0·95, so as to be on the safe side. We will consider only central intervals, so that for given ϖ we have to find p_0 and p_1 such that

$$P\ \{\varpi \geqslant p_0\} \geqslant 0{\cdot}975$$
$$P\ \{\varpi \leqslant p_1\} \geqslant 0{\cdot}975,$$

the inequalities for P being as near to equality as we can make them.

Consider the diagrammatic representation of the type shown in Fig. 19.1 and given for our present case in Fig. 19.2.

From the table we can find, for any assigned ϖ, the values ϖ_0 and ϖ_1 such that $P\ (p \geqslant \varpi_0) \geqslant 0{\cdot}975$ and $P\ (p \leqslant \varpi_1) \geqslant 0{\cdot}975$. Note that in determining ϖ_1 the distribution function gives the probability of obtaining a proportion p or less successes, so that the complement of the function gives the probability of a proportion $1 - p - 0{\cdot}05$ or less (not $1 - p$). Here, for example, on the horizontal through $\varpi = 0{\cdot}1$ we find $\varpi_0 = 0$ and $\varpi_1 = 0{\cdot}30$ from our table ; and for $\varpi = 0{\cdot}4$ we have $\varpi_0 = 0{\cdot}15$ and $\varpi_1 = 0{\cdot}65$. The points so obtained lie on stepped curves which have been drawn in. The zone between them is the confidence belt. For any p the probability that we shall be wrong in locating ϖ inside the belt is at the most 0·05. We determine p_0 and p_1 by drawing a vertical at the given value of p on the abscissa and reading off the values where it intersects the curves. That these are, in fact, the required limits will be shown in a moment.

We could have found more precise confidence limits by interpolating in the table obtained above. For example, with $p = 0{\cdot}30$ we see that

$$\text{for } \varpi = 0{\cdot}1,\ P = 0{\cdot}9977$$
$$\text{for } \varpi = 0{\cdot}2,\ P = 0{\cdot}9133.$$

Hence, for $P = 0{\cdot}975$ we have approximately

$$\varpi = 0{\cdot}1 + \frac{9977 - 9750}{9977 - 9133}(0{\cdot}1) = 0{\cdot}127,$$

and closer approximations can be obtained if desired. The corresponding point on the

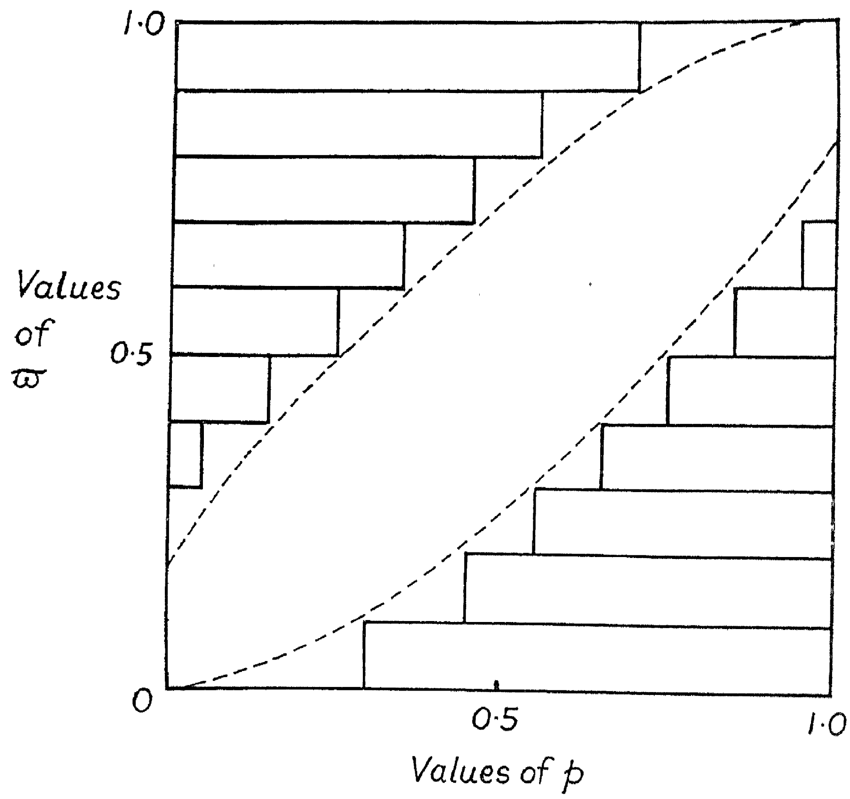


Fig. 19.2.

lower confidence line to $\varpi_1 = 0{\cdot}127$ is $p = 0{\cdot}35$. Calculations on these lines give us the values of ϖ such that

$$P\{p_0 \leqslant \varpi \leqslant p_1\} = \alpha \text{ exactly,}$$

whereas the former approach gave values such that

$$P\{p_0 \leqslant \varpi \leqslant p_1\} = \alpha \text{ approximately,}$$
$$\geqslant \alpha \text{ in any case.}$$

Discontinuous variates usually give rise to this sort of arithmetical nuisance, but the approximation in practice is sufficiently good, except for very small samples. The broken curves in Fig. 19.2 give the more precise limits. They lie, of course, inside the more approximate step-curves.

It is, perhaps, worth noticing that the points on the curves of Fig. 19.2 were constructed by selecting an ordinate ϖ and then finding the corresponding abscissae ϖ_0 and ϖ_1. The diagram is, so to speak, constructed *horizontally*. In applying it, however, we read it *vertically*, that is to say, with observed abscissa p we read off two values p_0 and p_1 and assert that $p_0 \leqslant \varpi \leqslant p_1$. It is instructive to observe how this change of viewpoint can be justified without reference to Bayes' postulate.

Consider Fig. 19.3, which shows a pair of confidence lines for the binomial. Let ϖ' be a given value of ϖ and let the horizontal through ϖ' meet the confidence lines in points with abscissae ϖ_0 and ϖ_1. Then we know that in repeated samples from a population with parameter ϖ' a proportion α will give observed values of p lying between ϖ_0 and ϖ_1; for the curves were constructed so that this should be so.

Now since the horizontal at ϖ' lies entirely within the confidence belt for $\varpi_0 \leqslant p \leqslant \varpi_1$ (and does so for any ϖ'), it follows that the assertion that ϖ' lies in the belt is correct if,

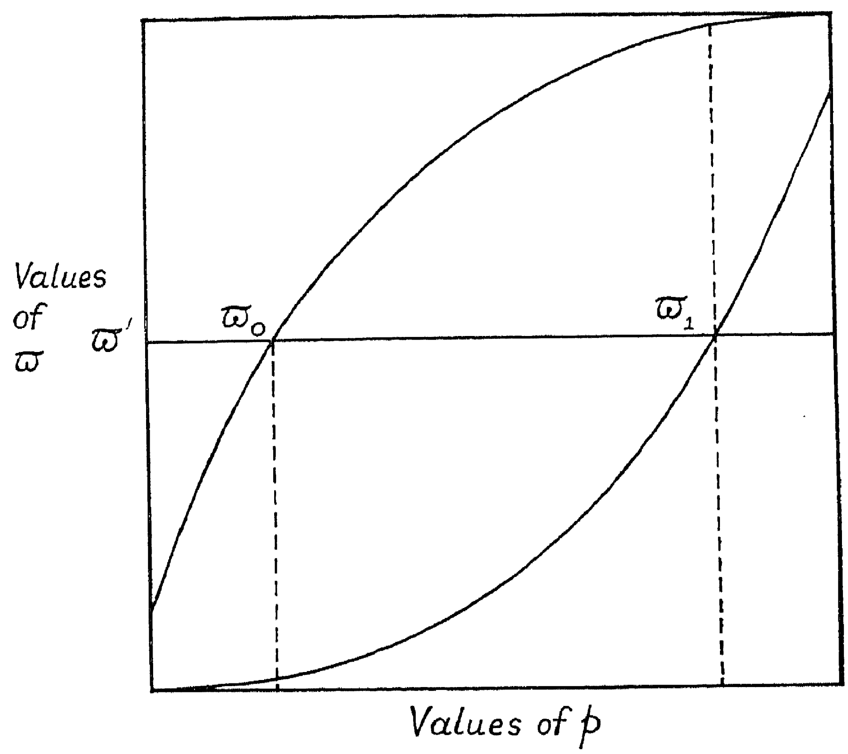


FIG. 19.3.

and only if, p lies between ϖ_0 and ϖ_1, that is in a proportion α of the cases. This, being true for any ϖ', is true for all ϖ', irrespective of the relative frequency of occurrence of the ϖ's under estimate. Consequently our assertion that ϖ lies in the confidence belt is correct in a proportion α of the cases; and, in particular, for any observed p we may assert that ϖ lies within the ordinates determined on the two curves by the vertical through p.

Confidence Intervals for Large Samples

19.10. In our usual notation, the logarithm of the likelihood function gives

$$\log L = \sum_{j=1}^{n} \log f(x_j, \theta), \qquad (19.5)$$

and

$$\frac{\partial \log L}{\partial \theta} = \Sigma \frac{\partial \log f}{\partial \theta}. \qquad (19.6)$$

We may regard $\dfrac{\partial \log L}{\partial \theta}$ as a random variable, and in particular write—

$$nA = \text{var}\left(\frac{\partial \log L}{\partial \theta}\right),$$

so that

$$A = \text{var}\left(\frac{\partial \log f}{\partial \theta}\right). \qquad (19.7)$$

Write
$$\psi = \frac{\dfrac{\partial \log L}{\partial \theta}}{\sqrt{(nA)}}. \qquad (19.8)$$

Then, for large samples, ψ will be distributed normally in the limit with unit variance, in virtue of the Central Limit Theorem, under very general conditions. It will also have zero mean, since

$$\begin{aligned} E\left(\frac{\partial \log f}{\partial \theta}\right) &= E\left(\frac{1}{f}\frac{\partial f}{\partial \theta}\right) \\ &= \int_{-\infty}^{\infty} \frac{\partial f}{\partial \theta}\, dx = \frac{\partial}{\partial \theta}\int f\, dx \\ &= \frac{\partial}{\partial \theta}.1 = 0. \end{aligned} \qquad (19.9)$$

Hence, from the distribution of ψ we may easily determine confidence limits for θ in large samples if ψ is a monotonic function of θ, so that inequalities in one may be transformed to inequalities in the other.

It is sufficient (but not necessary) for the existence of the normal limit to ψ that $\dfrac{\partial f}{\partial \theta}$ exists for all x, except perhaps at isolated points, that the range is independent of θ and that the Central Limit Theorem applies (e.g. if the third moment of $\dfrac{\partial \log f}{\partial \theta}$ exists). We also assume, as usual, that differentiation under the integral sign, as in (19.9), is legitimate.

Example 19.3

Consider again the problem of Example 19.1. We have, with μ for θ,

$$f(x, \mu) = \frac{1}{\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}(x-\mu)^2\right\}$$

$$\frac{\partial \log f}{\partial \mu} = x - \mu$$

$$\begin{aligned} \text{var}\left(\frac{\partial \log f}{\partial \mu}\right) &= \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} (x-\mu)^2 f\, dx \\ &= 1. \end{aligned}$$

Hence
$$\psi = \Sigma\left(\frac{x-\mu}{\sqrt{n}}\right) = (\bar{x} - \mu)\sqrt{n}$$

is normally distributed with unit variance for large n. (We know, of course, that this is true for small n as well in this particular case.) The confidence limits may then be set as in Example 19.1.

Example 19.4

Consider the Poisson distribution whose general term is

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x\,!}.$$

We have

$$\frac{\partial \log f}{\partial \lambda} = \frac{x}{\lambda} - 1$$

$$\operatorname{var}\left(\frac{\partial \log f}{\partial \lambda}\right) = \sum_{x=0}^{\infty}\left(\frac{x}{\lambda} - 1\right)^2 e^{-\lambda}\frac{\lambda^x}{x\,!}$$

$$= \frac{1}{\lambda}.$$

Hence

$$\psi = \frac{\frac{1}{\lambda}\Sigma(x) - n}{\sqrt{(n/\lambda)}} = \sqrt{\frac{n}{\lambda}}(\bar{x} - \lambda).$$

For example, with $\alpha = 0{\cdot}95$, corresponding to a normal deviate $\pm 1{\cdot}96$, we have, for the central confidence limits,

$$(\bar{x} - \lambda)\sqrt{\frac{n}{\lambda}} = \pm 1{\cdot}96,$$

giving, on solution for λ,

$$\lambda^2 - \left(2\bar{x} + \frac{3{\cdot}84}{n}\right)\lambda + \bar{x}^2 = 0$$

$$\lambda = \bar{x} + \frac{1{\cdot}92}{n} \pm \sqrt{\left(\frac{3{\cdot}84\bar{x}}{n} + \frac{3{\cdot}69}{n^2}\right)},$$

the ambiguity in the square root giving upper and lower limits respectively.

To order $n^{-\frac{1}{2}}$ this is equivalent to

$$\lambda = \bar{x} \pm 1{\cdot}96\sqrt{\frac{\bar{x}}{n}},$$

from which the upper and lower limits are seen to be equidistant from the mean $\bar{x}$, as we should expect.

Shortest Sets of Confidence Intervals

19.11. It has been seen in Example 19.1 that in some circumstances at least there exist more than one set of confidence intervals, and it is now necessary to consider whether any particular set can be regarded as better than the others in any useful sense. The problem is analogous to that of estimators, where we found that in general there are many different estimators for a parameter, but that we could sometimes find one (such as that with minimum variance) which was superior to the rest.

In Example 19.1 the problem presented itself in rather a specialised form. We found that for the intervals based on the mean $\bar{x}$ there were infinitely many sets of intervals according to the way in which we selected α_0 and α_1 (subject to the condition that $\alpha_0 + \alpha_1 = 1 + \alpha$). Among these the central intervals are obviously the shortest, for a given range will include the greatest area of the normal curve if it is centred at the mean of the curve. We might reasonably say that the central intervals are the best among those determined by $\bar{x}$.

But it does not follow that they are the shortest of all possible intervals, or even that such a shortest set exists. It might also happen that for two sets of intervals c_1 and c_2 those of c_1 are shorter than those of c_2 in part of the range of x's and longer in other parts.

19.12. We will therefore consider sets of intervals which are shortest on the average. That is to say, if

$$\delta = t_1 - t_0$$

we require that

$$\int \delta \, dF = \text{minimum}, \quad . \quad . \quad . \quad . \quad . \quad (19.10)$$

where the integral is taken over all x's and is therefore equivalent to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \delta \, L \, dx_1 \ldots dx_n. \quad . \quad . \quad . \quad . \quad (19.11)$$

We now prove a theorem which is very similar to the result that maximum-likelihood estimators in the limit have minimum variance, namely that in a certain class of intervals the method of **19.10** gives those which are shortest on the average.

Let $h\,(x, \theta)$ be a function which has a zero mean value and is such that the sum of a number of similar functions obeys the Central Limit Theorem. Then

$$\zeta = \frac{\sum_{j=1}^{n} h\,(x_j, \theta)}{\sqrt{(n \operatorname{var} h)}} \quad . \quad . \quad . \quad . \quad . \quad (19.12)$$

is normally distributed in the limit with zero mean and unit variance. ψ of equation (19.8) is a member of the class ζ. We prove that the average rate of change of ψ with respect to θ, for each fixed θ, is greater than that of any ζ except in the trivial case

$$h = k \frac{\partial \log f}{\partial \theta}.$$

Writing $g\,(x, \theta) = \dfrac{\partial \log f}{\partial \theta}$, we have

$$\frac{\partial \psi}{\partial \theta} = \frac{1}{\sqrt{(n \operatorname{var} g)}} \left\{ \Sigma \frac{\partial g}{\partial \theta} - \frac{1}{2 \operatorname{var} g} \Sigma g \frac{\partial \operatorname{var} g}{\partial \theta} \right\} \quad . \quad . \quad . \quad (19.13)$$

$$\frac{\partial \zeta}{\partial \theta} = \frac{1}{\sqrt{(n \operatorname{var} h)}} \left\{ \Sigma \frac{\partial h}{\partial \theta} - \frac{1}{\operatorname{var} h} \Sigma h \frac{\partial \operatorname{var} h}{\partial \theta} \right\} \quad . \quad . \quad . \quad (19.14)$$

Hence

$$E\left(\frac{\partial \psi}{\partial \theta}\right) = \frac{1}{\sqrt{(n \operatorname{var} g)}} \left\{ \Sigma E \left(\frac{\partial g}{\partial \theta}\right) - \frac{1}{2 \operatorname{var} g} \Sigma E\,(g) \frac{\partial \operatorname{var} g}{\partial \theta} \right\}.$$

Now $E\,(g) = 0$ and

$$E\left(\frac{\partial g}{\partial \theta}\right) = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right) = -E\left(\frac{\partial \log f}{\partial \theta}\right)^2$$
$$= -E\,(g^2).$$

Thus

$$E\left(\frac{\partial \psi}{\partial \theta}\right) = -\frac{nE\,(g^2)}{\sqrt{(n \operatorname{var} g)}}$$
$$= -\sqrt{(n \operatorname{var} g)} = \Delta_1, \text{ say}. \quad . \quad . \quad . \quad . \quad (19.15)$$

Similarly,

$$E\left(\frac{\partial \zeta}{\partial \theta}\right) = \sqrt{\frac{n}{\operatorname{var} h}}\, E\left(\frac{\partial h}{\partial \theta}\right) = \Delta_2, \text{ say}. \quad . \quad . \quad . \quad (19.16)$$

Since $E(h) = 0$ we have

$$E\left(\frac{\partial h}{\partial \theta}\right) = \int \frac{\partial h}{\partial \theta} f\, dx = -\int h \frac{\partial f}{\partial \theta} dx$$
$$= -\operatorname{cov}(h, g). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19.17)$$

Hence

$$\Delta_1^2 - \Delta_2^2 = n \operatorname{var} g - \frac{n}{\operatorname{var} h} \operatorname{cov}^2(h, g)$$
$$= \frac{n}{\operatorname{var} h}\{\operatorname{var} h \operatorname{var} g - \operatorname{cov}^2(h, g)\}. \quad . \quad . \quad (19.18)$$

Thus, unless h is a multiple of g, we have

$$\Delta_1^2 > \Delta_2^2,$$

which was to be proved.

Now if ψ_α is a value such that

$$\frac{1}{\sqrt{(2\pi)}} \int_0^{\psi_\alpha} e^{-\frac{1}{2}x^2}\, dx = \tfrac{1}{2}\alpha,$$

the upper and lower confidence points for central intervals are $\pm \psi_\alpha$ and the values of θ are the solutions of

$$\frac{\Sigma g(x, \theta)}{\sqrt{(n \operatorname{var} g)}} = \pm \psi_\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19.19)$$

say t_0 and t_1. Similarly those for any function h are given by

$$\frac{\Sigma h(x, \theta)}{\sqrt{(n \operatorname{var} h)}} = \pm \psi_\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19.20)$$

say u_0 and u_1. The equations for confidence points are equivalent to

$$\psi(t) = \pm \psi_\alpha$$
$$\zeta(u) = \pm \psi_\alpha$$

or, effectively, in large samples, by

$$\psi(\theta_0) + (t - \theta_0)\left(\frac{\partial \psi}{\partial \theta}\right)_{\theta_0} = \pm \psi_\alpha$$

$$\zeta(\theta_0) + (u - \theta_0)\left(\frac{\partial \zeta}{\partial \theta}\right)_{\theta_0} = \pm \psi_\alpha,$$

where θ_0 is a fixed value of θ. When $t = \theta_0$ and $u = \theta_0$ we have $\psi(\theta_0) = \zeta(\theta_0)$. Hence

$$(t - \theta_0)\left(\frac{\partial \psi}{\partial \theta}\right)_{\theta_0} = (u - \theta_0)\left(\frac{\partial \zeta}{\partial \theta}\right)_{\theta_0}. \quad . \quad . \quad . \quad . \quad (19.21)$$

Now we have just shown that, on the average, $\dfrac{\partial \psi}{\partial \theta} > \dfrac{\partial \zeta}{\partial \theta}$. Hence, on the average,

$$t - \theta_0 < u - \theta_0,$$

and the confidence limits t are closer together than those of any member of the class u for any fixed value of θ.

19.13. A comparison of the result we have just proved and the properties of maximum likelihood estimators in the limit will show the close relation between confidence intervals and the theory of estimation developed in Chapter 17. In **17.27** we showed,

by considering the quantity $u = \frac{\partial \log L}{\partial \theta}$, that any estimator t which is in the limit distributed normally about the true value θ_0 cannot have a variance less than

$$1/n\, E\left(\frac{\partial \log f}{\partial \theta}\right)^2;$$

and that the latter quantity, in the limit, is the variance of the maximum likelihood estimator. It attains the minimal value when u is constant over samples for which t is constant.

The theorem of **19.12** shows that on the average the intervals determined by the distribution of u are shorter than those based on any other function with a zero mean value (obeying the usual conditions as to continuity, etc.). Since the maximum likelihood estimator has minimum variance, we should expect that confidence intervals based on its distribution would be shorter than others; and this we now see to be so. For if u is constant over samples of constant t, the distribution of u in all samples is equivalent to that of t.

Confidence Intervals and Sufficient Estimators

19.14. Pursuing this line of thought, we are led to inquire whether sufficient estimators provide confidence intervals for finite samples and whether they have any minimal properties of the kind we have just established for large samples.

It is easy to see that sufficient estimators do in fact provide confidence intervals. If t is sufficient for θ, the likelihood function may be put in the form

$$L = f_1(t, \theta)\, f_2(x_1 \ldots x_n) \qquad (19.22)$$

and the distribution of t and θ is

$$dF = f_1(t, \theta)\, dt. \qquad (19.23)$$

Given α we can then find t_0 and t_1 such that $F(t_0, \theta) = 1 - \alpha_0$ and $F(t_1, \theta) = \alpha_1$ and solve for θ in terms of t_0 and α_0 or t_1 and α_1, as the case may be. This process will provide the inequalities of the type we require, a proposition which we shall prove formally below **(19.25)**.

Example 19.5

In Example 17.8 we saw that

$$\hat{\theta} = \frac{\bar{x}}{p}$$

is sufficient for θ in the distribution

$$dF = \frac{x^{p-1} e^{-x/\theta}}{\Gamma(p)\, \theta^p}\, dx, \qquad 0 \leqslant x \leqslant \infty, \quad p > 1,$$

where p is regarded as known. The distribution of $\hat{\theta}$ is in fact

$$dF = \left(\frac{np}{\theta}\right)^{np} \frac{\hat{\theta}^{np-1} \exp\left(-\frac{np\,\hat{\theta}}{\theta}\right)}{\Gamma(np)}\, d\hat{\theta}.$$

The distribution function of $m = \frac{np\,\hat{\theta}}{\theta}$ is the incomplete Γ-function

$$\frac{\Gamma_m(np)}{\Gamma(np)} = I\left(\frac{m}{\sqrt{(np)}},\ np - 1\right).$$

We then find the values of m corresponding to α_0 and α_1 from the tables, and have

$$P\,(m \leqslant m_0) = \alpha_0$$
$$P\,(m \geqslant m_1) = \alpha_1,$$

whence

$$P\left\{\frac{np\hat{\theta}}{m_0} \leqslant \theta \leqslant \frac{np\hat{\theta}}{m_1}\right\} = \alpha_0 + \alpha_1 - 1$$
$$= \alpha.$$

19.15. The position in regard to minimal properties of confidence intervals based on sufficient estimators remains somewhat obscure, but one would expect some such properties to hold even for finite n. Since $u = \dfrac{\partial \log L}{\partial \theta}$ is constant for constant t when t is sufficient, the variance of u will be a function of the variance of t. This, however, is not necessarily enough to establish the fact that the corresponding confidence intervals are shortest on the average. It is imaginable that the confidence intervals derived from its distribution might be longer on the average than those of some other system. This seems rather unlikely, at least for the ordinary distributions of statistical theory, but apparently no proof has been given.

19.16. Neyman (1937*b*) has proposed to apply the phrase "shortest confidence intervals" to sets of intervals defined in quite a different way. As it does not appear that such intervals are necessarily the shortest in the sense of possessing the least length, even on the average, we shall attempt to avoid confusion by calling them "most selective."

Consider a set of intervals c_0, typified by δ, obeying the condition that

$$P\,\{\delta_0\, c\, \theta \mid \theta\} = \alpha, \qquad . \quad . \quad . \quad . \quad . \quad (19.24)$$

where we write $\delta_0\, c\, \theta$—that is, δ_0 "contains" θ—for the more usual $t_0 \leqslant \theta \leqslant t_1$ $(t_1 - t_0 = \delta_0)$. Let c_1 be some other set typified by δ_1 such that

$$P\,\{\delta_1\, c\, \theta \mid \theta\} = \alpha. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (19.25)$$

Either set is a permissible set of intervals, as the probability is α in both cases that the range δ contains θ.

If now for every c_1 we have, for any value θ' other than the true value,

$$P\,\{\delta_0\, c\, \theta' \mid \theta\} \leqslant P\,\{\delta_1\, c\, \theta' \mid \theta\}, \qquad . \quad . \quad . \quad . \quad (19.26)$$

c_0 is said to be *most selective.*

19.17. The ideas underlying this definition will be clearer from a reading of Chapters 26 and 27 dealing with the Neyman-Pearson theory of inference. We anticipate them here to the extent of remarking that the object of most selective intervals is to cover the true value with assigned probability α, but to cover other values as little as possible. We may say of both c_0 and c_1 that the assertion $\delta\, c\, \theta$ is true in proportion α of the cases. What marks out c_0 for choice as the most selective set is that it covers false values less frequently than the remaining sets.

The difference between this approach and the one leading to shortest intervals is that the latter is concerned only with the narrowness of the confidence interval, whereas the former gives weight to the frequency with which alternative values of θ are covered. One

concentrates on locating θ with the smallest margin of error; the other takes into account the desirability of excluding so far as possible false values of θ from the interval, so that mistakes of taking the wrong value are minimised.

19.18. Neyman himself has shown that most selective sets do not usually exist (for instance, if the distribution is continuous) and has proposed two alternative systems :—

(*a*) most selective one-sided systems (Neyman's "shortest one-sided" sets) which obey (19.26) only for values of $\theta' - \theta$ which are always positive or always negative;

(*b*) selective unbiassed systems (Neyman's "short unbiassed" sets) which obey (19.25) but, in place of (19.26), the further relation

$$P\{\delta \, c \, \theta \mid \theta\} = \alpha \geqslant P\{\delta \, c \, \theta \mid \theta'\}. \qquad (19.27)$$

In essence these sets amount to a translation into terms of confidence intervals of certain ideas in the theory of tests of significance, and we may defer consideration of them until Chapters 26 and 27 are reached.

Generalisation to the Case of Several Parameters

19.19. We now proceed to generalise the foregoing theory to the case of several parameters. Although, to simplify the exposition, we shall deal in detail only with a single variate, the theory is quite general. We begin by extending our notation and introducing a geometrical terminology which may be regarded as an elaboration of the diagrams of Figs. 19.1 and 19.2.

Suppose we have a frequency function of known form depending on l unknown parameters, $\theta_1 \ldots \theta_l$, and denoted by $f(x, \theta_1 \ldots \theta_l)$. We may require to estimate either θ_1 only or several of the θ's simultaneously. In the first place we consider only the estimation of a single parameter. To determine confidence limits we require to find two functions u_0 and u_1, dependent on the sample values but not on the θ's, such that

$$P\{u_0 \leqslant \theta_1 \leqslant u_1 \mid \theta_1 \ldots \theta_l\} = \alpha, \qquad (19.28)$$

where α is the confidence coefficient chosen in advance.

With a sample of n values, $x_1 \ldots x_n$, we can associate a point in an n-dimensional Euclidean space, and the frequency-distribution will determine a density function for each such point. The quantities u_0 and u_1, being functions of the x's, are determined in this space, and for any given α will lie on two hypersurfaces (the natural extension of the confidence lines of Fig. 19.1). Between them will lie a Confidence Zone or Region of Acceptance.

In general we also have to consider a range of values of θ which are *a priori* possible. There will thus be an l-dimensional space of θ's subjoined to the n-space, the total region of variation having $(l + n)$ dimensions; but if we are considering the estimation of θ_1, this reduces to an $(n + 1)$-space, the other $(l - 1)$ parameters not appearing as variables.

We shall call the sample-space W and denote a point whose co-ordinates are $x_1 \ldots x_n$ by E. We may then write $u_0(E)$, $u_1(E)$ to show that the confidence functions depend on E. The interval $u_1(E) - u_0(E)$ we denote by $\delta(E)$ or δ, and as above we write $\delta \, c \, \theta_1$ to denote $u_0 \leqslant \theta_1 \leqslant u_1$. The region of acceptance or confidence zone we denote by A, and may write $E \, \varepsilon \, \delta$ or $E \, \varepsilon \, A$ to indicate that the sample-point lies in the interval δ or the region A.

19.20. In Fig. 19.4 we have shown two axes x_1 and x_2 and a third axis corresponding to the variation of θ_1. The sample-space W is thus two-dimensional. For any given θ_1, say θ_1', the space W is a hyperplane (or part of it), one such being shown.

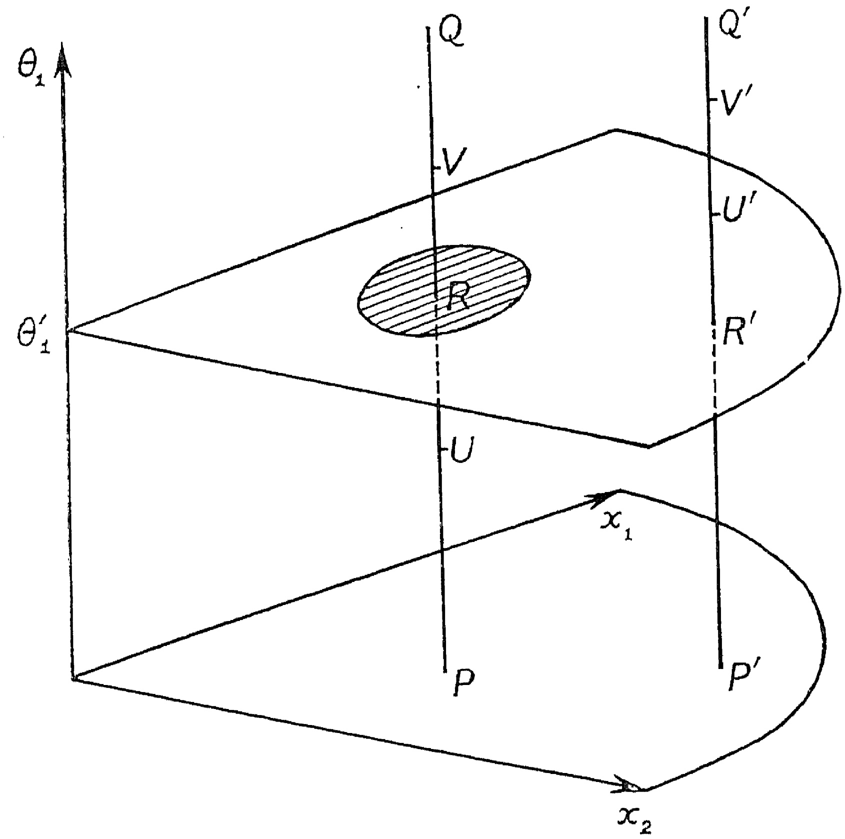


FIG. 19.4.

Take any given pair of values (x_1, x_2) and draw through the point so defined a line parallel to the θ_1-axis, such as PQ in the figure, cutting the hyperplane at R. The two values of u_0 and u_1 will give two limits to θ_1 corresponding to two points on this line, say U, V. Consider now the lines PQ as x_1, x_2 vary. In some cases U, V will lie on opposite sides of R, and θ_1' lies inside the interval UV. In other cases (as for instance in $U'V'$ shown in the figure) the contrary is true. The totality of points in the former category determines the region of acceptance A, shaded in the figure. If for any point in A we assert $\delta\ c\ \theta_1'$, we shall be right; if we assert it for points outside A we shall be wrong.

19.21. Evidently, if the sample-point E falls in the region A, the corresponding θ_1' lies in the confidence interval and conversely. It follows that the probability of any *fixed* θ_1' lying in the confidence interval is the probability that E lies in $A\,(\theta_1')$; or in symbols—

$$P\{\delta\ c\ \theta_1' \mid \theta_1 \ldots \theta_l\} = P\{u_0 \leqslant \theta_1' \leqslant u_1 \mid \theta_1 \ldots \theta_l\}$$
$$= P\{E\ \varepsilon\ A\,(\theta_1') \mid \theta_1 \ldots \theta_l\}. \qquad (19.29)$$

From this it follows that if the confidence functions are determined so that

$$P\{u_0 \leqslant \theta_1 \leqslant u_1 \mid \theta_1 \ldots \theta_l\} = \alpha$$

we shall have, for all θ_1,

$$P\{E\ \varepsilon\ \mathrm{A}\,(\theta_1) \mid \theta_1 \ldots \theta_l\} = \alpha. \qquad (19.30)$$

It follows also that for no θ_1 can the region A be empty, for if it were the probability in (19.30) would be zero.

19.22. If the functions u_0 and u_1 are single-valued and determined for all E, then any sample-point will fall into at least one region of acceptance. For on the line PQ corresponding to the given E we take an R between U and V, and this will define a value of θ_1, say θ_1', such that $E \varepsilon A(\theta_1')$.

More importantly, if a sample-point falls in the regions $A(\theta_1')$ and $A(\theta_1'')$ corresponding to two values of θ_1, θ_1' and θ_1'', it will fall in the region $A(\theta_1''')$, where θ_1''' is any value between θ_1' and θ_1''. For we have

$$u_0 \leqslant \theta_1' \leqslant u_1, \qquad u_0 \leqslant \theta_1'' \leqslant u_1,$$

and hence

$$u_0 \leqslant \theta_1' \leqslant \theta_1'' \leqslant u_1$$

if θ_1'' is the greater, and hence

$$u_0 \leqslant \theta_1' \leqslant \theta_1''' \leqslant \theta_1'' \leqslant u_1$$

or

$$u_0 \leqslant \theta_1''' \leqslant u_1.$$

Further, if a sample-point falls in any of the regions $A(\theta_1)$ for the range of θ-values $\theta_1' < \theta_1 < \theta_1''$, it must also fall within $A(\theta_1')$ and $A(\theta_1'')$.

19.23. The conditions referred to in the two previous sections are necessary. We now prove that they are sufficient, that is to say: if for each value of θ_1 there is defined in the sample-space W a region A such that

(1) $P\{E \varepsilon \mathrm{A}(\theta_1) \mid \theta_1\} = \alpha$, whatever the value of the θ's;

(2) For any E there is at least one θ_1, say θ_1', such that $E \varepsilon \mathrm{A}(\theta_1')$;

(3) If $E \varepsilon A(\theta_1')$ and $E \varepsilon A(\theta_1'')$, then $E \varepsilon A(\theta_1''')$ for any θ_1''' between θ_1' and θ_1'';

(4) If $E \varepsilon A(\theta_1)$ for any θ_1 satisfying $\theta_1' < \theta_1 < \theta_1''$, $E \varepsilon A(\theta_1')$ and $E \varepsilon A(\theta_1'')$;

then u_0 and u_1, viz. confidence limits for θ, are given by taking the lower and upper bounds of values of θ_1 for which a fixed sample-point falls within $A(\theta_1)$. They are determinate and single-valued for all E, $u_0 \leqslant u_1$, and $P\{u_0 \leqslant \theta_1 \leqslant u_1 \mid \theta_1\} = \alpha$ for all θ_1.

The lower and upper bounds exist in virtue of condition (2), and the lower is not greater than the upper. We have then merely to show that $P\{u_0 \leqslant \theta_1 \leqslant u_1 \mid \theta_1\} = \alpha$, and for this it is sufficient, in virtue of condition (1), to show that

$$P\{u_0 \leqslant \theta_1 \leqslant u_1 \mid \theta_1\} = P\{E \varepsilon A(\theta_1) \mid \theta_1\}. \quad . \quad . \quad (19.31)$$

We already know that if $E \varepsilon A(\theta_1)$ then $u_0 \leqslant \theta_1 \leqslant u_1$; and our result will be established if we demonstrate the converse.

Suppose it is not true that when $u_0 \leqslant \theta_1 \leqslant u_1$, $E \varepsilon A(\theta_1)$. Let E' be a point outside $A(\theta_1)$ for which $u_0 \leqslant \theta_1 \leqslant u_1$. Then must either $u_0 = \theta_1$ or $u_1 = \theta_1$ or both; for otherwise u_0 and u_1 being the bounds of the values of θ_1 for which E lies in $A(\theta_1)$, there would exist values θ_1' and θ_1'', such that $E \varepsilon A(\theta_1')$ and $E \varepsilon A(\theta_1'')$ and

$$u_0 \leqslant \theta_1' < \theta_1 < \theta_1'' \leqslant u_1,$$

so that, from condition (3), $E \varepsilon A(\theta_1)$ which is contrary to assumption.

Thus $u_0 = \theta_1$ or $u_1 = \theta_1$ or both. If both, then E must fall in $A(\theta_1)$, for u_0 and u_1 are the bounds of θ-values for which this is so, and if they coincide their common value must be so. Finally, if $u_0 = \theta_1 < u_1$ (and similarly if $u_0 < \theta_1 = u_1$) we see that for $u_0 < \theta_1 < u_1$, E must fall in $A(\theta_1)$ from condition (3), and hence, from condition (4), E must fall in $A(\theta_1')$ and $A(\theta_1'')$ where $\theta_1' = u_0$ and $\theta_1'' = u_1$. Hence it falls in $A(\theta_1)$.

19.24. The foregoing theorem gives us a formal solution of the problem of finding confidence intervals in the general case, but it does not provide a method of finding the

intervals in particular instances. In practice we have three lines of approach: (1) to use sufficient estimators, (2) to adopt the process known as "studentisation," and (3) to "guess" a set of intervals in the light of general knowledge and experience and to verify that they do or do not satisfy the required conditions.

19.25. Consider the use of sufficient estimators in the general case. If t_1 is sufficient for θ_1 we have

$$L = L_1(t_1, \theta_1)\, L_2(x_1 \ldots x_n, \theta_2 \ldots \theta_l). \qquad (19.32)$$

The locus t_1 = constant determines a series of hypersurfaces in the sample-space W. If we regard these hypersurfaces as determining regions in W, then $t_1 \leqslant k$, say, determines a fixed region K. The probability that E falls in K is then clearly dependent only on t_1 and θ_1. By appropriate choice of k we can determine K so that

$$P\{E \,\varepsilon\, K \mid \theta_1\} = \alpha,$$

and hence set up regions of acceptance based on values of t_1. We can do so, moreover, in an infinity of ways, according to the values selected for α_0 and α_1.

Studentisation

19.26. In Example 19.1 we considered a simplified problem of estimating the mean in samples from a normal population with unit variance. Suppose now that we require to determine confidence limits for the mean μ in samples from

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx.$$

The approach of Example 19.1 would lead us to the conclusion that, for confidence coefficient 0·9545 and central intervals,

$$P\left\{\bar{x} - \frac{2\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + \frac{2\sigma}{\sqrt{n}} \mid \mu, \sigma\right\} = 0{\cdot}9545.$$

But we cannot now say that the confidence limits are $\bar{x} \pm 2\sigma/\sqrt{n}$ because σ is unknown.

Consider then the distribution of $z = \dfrac{\bar{x}-\mu}{s}$, where s^2 is the sample variance. This is known to be the "Student" form

$$dF = \frac{k\,dz}{(1+z^2)^{\frac{n}{2}}}.$$

(Cf. Example 10.6, vol. I, p. 239.) Given α, we can now find z_0 and z_1, such that

$$\int_{-\infty}^{-z_1} dF = \int_{z_0}^{\infty} dF = \frac{1-\alpha}{2},$$

and hence

$$P\{-z_1 \leqslant z \leqslant z_0\} = \alpha,$$

which is equivalent to

$$P\{\bar{x} - sz_0 \leqslant \mu \leqslant \bar{x} + sz_1\} = \alpha.$$

Hence we may say that μ lies in the range $\bar{x} - sz_0$ to $\bar{x} + sz_1$ with confidence coefficient α, the range now being independent of either μ or σ. In fact, owing to the symmetry of "Student's" distribution, $z_0 = z_1$, but this is an accidental circumstance peculiar to the present case.

19.27. The possibility of finding confidence intervals in this case arose from our being able to find a statistic z, depending only on the parameter under estimate, whose distribution did not contain σ. A scale parameter can often be eliminated in this way, although the resulting distributions are not always easy to handle. If, for instance, we have a statistic t which is of degree p in the variables, then t/s^p is of degree zero, and its distribution must be independent of the scale parameter. When a statistic is reduced to independence of the scale in this way it is said to be "studentised," after "Student" (W. S. Gosset), who was the first to perceive the significance of the process.

19.28. It is interesting to consider the relation between the studentised mean-statistic and confidence zones based on sufficient estimators in the normal case. The distribution of means and variances in normal samples is

$$dF = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} d\bar{x} \frac{k}{\sigma^{n-\frac{1}{2}}} s^{n-3} \exp\left(-\frac{ns^2}{2\sigma^2}\right) ds^2 \quad . \quad (19.33)$$

and $\bar{x}$, s are jointly sufficient for μ, σ. In the sample space W the regions of constant $\bar{x}$ are hyperplanes and those of constant s are hyperspheres. If we fix $\bar{x}$ and s the sample-point E lies on a hypersphere of $(n-2)$ dimensions. Choose an area on this hypersphere of content α. Then the acceptance region will be obtained by combining all such areas for all $\bar{x}$ and s.

One such region is seen to be the "slice" of the sample-space obtained by rotating the hyperplane passing through the origin and the point $(1, 1 \ . \ . \ . \ 1)$ through an angle $\pi\alpha$ (not $2\pi\alpha$ because a half-turn of the plane covers the whole space).

The situation is illustrated for $n = 2$ in Fig. 19.5.

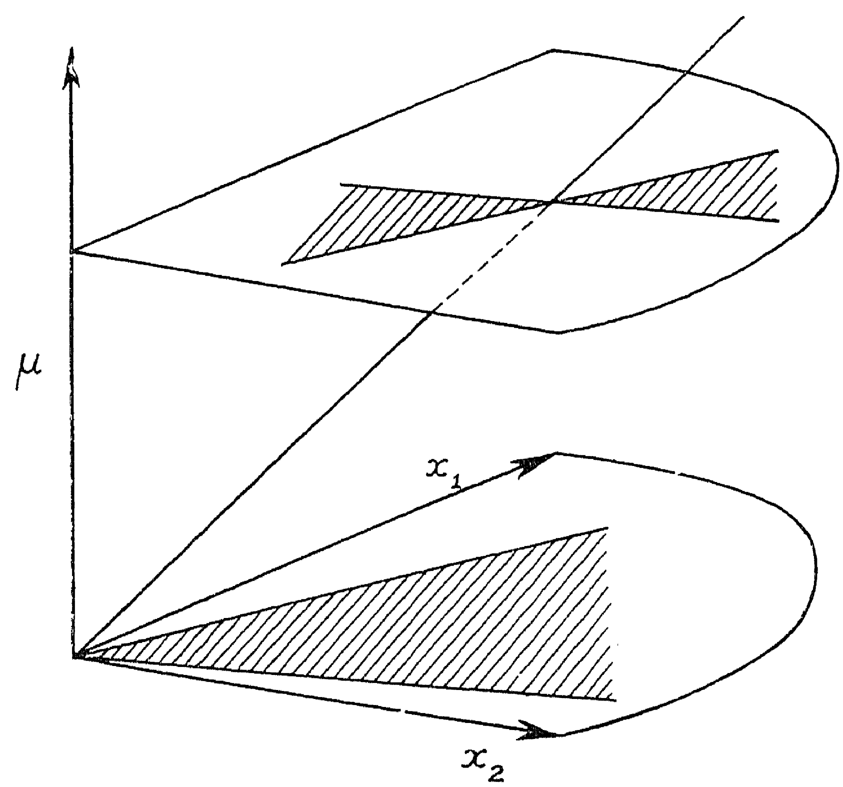


FIG. 19.5.

For any given μ' the axis of rotation meets the hyperplane $\mu = \mu'$ in the point $x_1 = x_2 = \mu'$, and the hypercones $\dfrac{\bar{x}-\mu}{s} =$ constant in the W space become the plane

areas between two straight lines (shaded in the figure). These may be regarded as regions of acceptance, and one set is that obtained by rotating a plane about the line $x_1 = x_2 = \mu$ through an angle so as to cut off in any plane $\mu = \mu'$ an angle $\frac{\pi\alpha}{2}$ on each side of

$$x_1 - \mu' = x_2 - \mu'.$$

The boundary planes are given by

$$x_1 - \mu = (x_2 - \mu) \tan\left(\frac{\pi}{4} - \frac{\beta}{2}\right)$$

$$x_1 - \mu = (x_2 - \mu) \tan\left(\frac{\pi}{4} + \frac{\beta}{2}\right),$$

where $\beta = \pi(1 - \alpha)$; or, after a little reduction,

$$\mu = \frac{x_1 + x_2}{2} + \frac{x_1 - x_2}{2} \cot\frac{\beta}{2}$$

$$\mu = \frac{x_1 + x_2}{2} - \frac{x_1 - x_2}{2} \cot\frac{\beta}{2}.$$

μ then lies in the region of acceptance if

$$\frac{x_1 + x_2}{2} - \frac{|x_1 - x_2|}{2} \cot\frac{\beta}{2} \leqslant \mu \leqslant \frac{x_1 + x_2}{2} + \frac{|x_1 - x_2|}{2} \cot\frac{\beta}{2}.$$

These are in fact the limits given by "Student's" distribution for $n = 2$, since the sample variance then becomes $\left|\frac{x_1 - x_2}{2}\right|^2$ and

$$\frac{1}{\pi}\int_{z_0}^{\infty} \frac{dz}{1 + z^2} = \frac{1}{\pi}\left(\frac{\pi}{2} - \tan^{-1} z_0\right) = \frac{1 - \alpha}{2} = \frac{\beta}{2\pi},$$

so that

$$z_0 = \tan\left(\frac{\pi}{2} - \frac{\beta}{2}\right) = \cot\frac{\beta}{2}.$$

19.29. Tables or diagrams of the confidence intervals for selected values of α have been given for the following parameters :—

(*a*) the proportion ϖ in the binomial (Clopper and Pearson, 1934);
(*b*) the parameter of the Poisson distribution (Garwood, 1936; Ricker, 1937);
(*c*) the correlation coefficient in normal samples (David, 1938*a*);
(*d*) the median in samples from any population (K. R. Nair, 1940*b*).

In addition, results for the mean of a normal population may be obtained from "Student's" integral as shown above. Those for the variance of a normal population may be obtained from the Γ-function or the equivalent χ^2-integral. For simultaneous estimation of mean and variance there are difficulties, as we proceed to show.

19.30. It might have been expected that the foregoing theory could be generalised to give simultaneous pairs of confidence intervals for two unknown parameters when intervals for each separately cannot be found. Very little progress in this direction has, however, been made. The difficulty may be illustrated by reference to the joint distri-

bution of mean and variance (19.33). From the independent distributions of $\bar{x} - \mu$ and $\frac{s}{\sigma}$ we can, given α, β, find t_0, t_1 and u_0, u_1 such that

$$P\left\{-t_1 \leqslant \frac{\bar{x}-\mu}{\sigma} \leqslant t_0\right\} = \alpha$$

$$P\left\{u_0 \leqslant \frac{s}{\sigma} \leqslant u_1\right\} = \beta$$

where the t's and u's depend only on sample values and α, β may be chosen at will. The *inequalities* are equivalent to

$$\bar{x} - \sigma t_0 \leqslant \mu \leqslant \bar{x} + \sigma t_1 \quad . \quad . \quad . \quad . \quad . \quad (19.34)$$

$$\frac{s}{u_1} \leqslant \sigma \leqslant \frac{s}{u_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19.35)$$

and these give

$$\bar{x} - \frac{t_0}{u_0} s \leqslant \mu \leqslant \bar{x} + \frac{t_1}{u_1} s. \quad . \quad . \quad . \quad . \quad (19.36)$$

But can we then infer that

$$P\left\{\bar{x} - \frac{t_0}{u_0} s \leqslant \mu \leqslant \bar{x} + \frac{t_1}{u_1} s\right\} = \gamma, \quad . \quad . \quad . \quad (19.37)$$

where γ is a constant dependent on α and β ? We cannot. This equation is, in fact, not generally true. The fact can be verified by considering the distribution of the statistic $\bar{x} - ks$ and showing that its distribution function $F(u)$ is not independent of μ and σ.

19.31. In the next chapter we shall see that a similar problem, giving rise to Behrens' test, provides a crucial point of difference between the theory of confidence intervals and that of fiducial intervals. All we need say here is that from the point of view of the former the problem of simultaneous confidence intervals for several parameters remains unsolved, except of course in the degenerate case when we can find independent intervals for each parameter separately.

19.32. In conclusion we indicate without proof a few results which have recently been obtained.

(1) Wilks and Daly (1939*b*) have generalised the theorem of **19.12** to the case of several parameters. Under fairly general conditions the confidence regions which are shortest on the average are given by

$$\frac{1}{n} \Sigma \left\{a_{ij} \frac{\partial \log L}{\partial \theta_i} \frac{\partial \log L}{\partial \theta_j}\right\} \leqslant \chi^2_\alpha,$$

where (a_{ij}) is the inverse matrix to that whose general element is

$$E\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right)$$

and χ^2_α is such that $P(\chi^2 \leqslant \chi^2_\alpha) = \alpha$, the probability being calculated from the χ^2-distribution with $\nu = l$. This is clearly related to the result of **17.46** giving the limiting forms of variances and covariances of maximum likelihood estimators.

(2) Wald (1942) has considered the problem of large samples from the point of view of most selective sets ("shortest" in Neyman's sense) and has proved results somewhat similar to those of Wilks and Daly.

(3) Wald and Wolfowitz (1939*b*, 1941*c*) and Kolmogoroff (1941) have considered the problem of setting confidence limits to the terminals of an unknown frequency-distribution.

NOTES AND REFERENCES

When the theory of confidence intervals and that of fiducial intervals were first developed many statisticians regarded them as equivalent. In papers written between 1930 and 1938 "confidence limits" and "fiducial limits" are often used in the same sense; and even where a distinction of approach was drawn the results given by the two methods appeared identical. The case of Behrens' test, however, provided an illustration where the methods lead to different results—see the following chapter.

The fiducial approach is due to R. A. Fisher, references being given at the end of Chapter 20. The approach of the present chapter has been developed mainly by Neyman (see particularly 1937*b*), E. S. Pearson, Wilks (1938*b*, *c*, 1939*a* and—with Daly—1939*b*), Wald (1939*a*, 1942), Welch (1939*a*), and Bartlett (1936*a*, 1939*a*). A number of the references to Chapters 26 and 27 are also relevant.

Confidence intervals can be obtained for the median and other quantiles which are independent of the form of distribution. See Thompson (1936), Savur (1937*a*) and K. R. Nair (1940*b*), and compare Exercise 19.5.

EXERCISES

19.1. Show that for the rectangular population

$$dF = \frac{dx}{\theta}, \qquad 0 \leqslant x \leqslant \theta$$

and confidence coefficient α, confidence limits for θ are t and t/ψ where t is the sample range and ψ is given by

$$\psi^{n-1}\{n - (n-1)\psi\} = 1 - \alpha.$$

(Wilks, 1938*c*.)

19.2. Show that, for the distribution of the previous exercise, confidence limits for samples of two, x_1 and x_2, are

$$\frac{x_1 + x_2}{1 + \sqrt{(1-\alpha)}}, \qquad \frac{x_1 + x_2}{1 - \sqrt{(1-\alpha)}}.$$

(Neyman, 1937*b*.)

19.3. Show also, in the case of the previous exercises, that if L is the larger of a sample of two, confidence limits are

$$L, \qquad \frac{L}{\sqrt{(1-\alpha)}}.$$

(Neyman, 1937*b*.)

Show further that if M is the largest of samples of four, confidence limits are

$$M, \qquad \frac{M}{(1-\alpha)^{\frac{1}{4}}}.$$

(For an experimental verification, see Frankel and Kullback, 1940.)

19.4. Show that, for the distribution

$$dF = \theta\, e^{-x\theta}\, dx, \qquad 0 \leqslant x \leqslant \infty$$

central confidence limits for large samples with $\alpha = 0{\cdot}95$ are given by

$$\theta = \frac{1 \pm \dfrac{1{\cdot}96}{\sqrt{n}}}{\bar{x}}.$$

(Wilks, 1938*c*.)

19.5. If a frequency function is continuous, the probability that the kth of a sample of n (arranged in ascending order of magnitude) lies in the range dx is

$$\frac{1}{B\,(k,\, n - k + 1)} F^{k-1}\,(1 - F)^{n-k}\, dF,$$

where F is the distribution function. Deduce that

$$P\,\{x_k < M < x_{n-k+1}\} = 1 - 2\, I_{0\cdot5}\,(n - k + 1,\, k),$$

where M is the median, and hence show how to determine confidence intervals for M from the incomplete B-function.

Generalise the result for quantiles. Show that the results do not hold for discontinuous distributions.

(Thompson, 1936.)

CHAPTER 20

FIDUCIAL INFERENCE

20.1. We now proceed to examine a type of inference known as fiducial. As in other methods of estimation, given a distribution of known form depending on an unknown parameter θ, we shall attempt to find limits between which θ lies in some sense associated with the theory of probability. To that extent our present approach is similar to the use of estimators with their associated sampling error and to the use of confidence intervals; but it is distinct from the latter both in essential ideas and in some of the results to which it leads.

20.2. Consider samples of n from a normal population of unknown mean μ and unit variance. The sample-mean $\bar{x}$ is sufficient for μ and its distribution is

$$dF = \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\} d\bar{x}. \quad \quad (20.1)$$

In speaking of a distribution in this sense we regard μ as fixed and consider the totality of values of $\bar{x}$ derived by random sampling from the population with given μ. The proportion of samples falling in a range $d\bar{x}$ is then given by (20.1), which holds for each value of μ.

We now change our viewpoint and consider a different kind of distribution based on (20.1). If we are given a value of $\bar{x}$ from a sample, what are the values of μ which could have given rise to this value to any fixed level of probability? If the deviation $\bar{x} - \mu$ is written as h, we know that the probability of the inequality

$$\bar{x} - \mu \leqslant h \quad \quad (20.2)$$

being true is α, where α depends on h and is in fact

$$\int_{-\infty}^{h} \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nx^2}{2}\right) dx. \quad \quad (20.3)$$

Looking at this the other way round, we may say that given any α we can find h, a function of α only, such that

$$\mu \geqslant \bar{x} - h \quad \quad (20.4)$$

is true with probability α. For any *fixed* $\bar{x}$ this gives us a distribution of μ. Consider in fact the equation

$$\mu = \bar{x} - h. \quad \quad (20.5)$$

If μ has a distribution function $F(\mu)$, we have, since (20.4) is true with probability α,

$$1 - \alpha = F(\mu) = 1 - \int_{-\infty}^{h} \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nx^2}{2}\right) dx,$$

whence

$$f(\mu)\, d\mu = -\sqrt{\frac{n}{2\pi}} \exp\left(-\frac{nh^2}{2}\right) dh.$$

But in virtue of (20.5), $d\mu = -dh$ and $h = \mu - \bar{x}$. Thus

$$f(\mu)\, d\mu = \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{n(\mu-\bar{x})^2}{2}\right) d\mu. \quad \quad (20.6)$$

This is called the *fiducial* distribution of μ.

20.3. It so happens that in this example the non-differential parts of (20.6) and (20.1) are the same. This is not essential although it is not infrequent. The crucial point of difference, however, lies in the appearance of the differential element $d\mu$, relating to the variation of μ, and the disappearance of $d\bar{x}$ relating to the variation of $\bar{x}$. We have derived a distribution of the parameter μ from that of the random variable $\bar{x}$ by transferring our attention in (20.4) from $\bar{x}$ to μ and regarding the inequality as still satisfied with probability α.

20.4. We note in the first place that this distribution is not necessarily existent. When we come to make an inference in any particular case we do not assume that μ is itself distributed in the fiducial form in the sense that it has been chosen at random from an existent population of μ's of that form. Such a prior distribution, which would be required for the application of Bayes' theorem, is not admissible from the point of view of the frequency theory of probability. The fiducial distribution is a hypothetical one of *conceivable* values of μ. We attach probabilities to these values, or rather to values in the range $d\mu$, by identifying them with the probabilities (based on frequency) which are derived from the distribution of a sufficient estimator of μ. For this reason the fiducial distribution is not a frequency-distribution in the ordinary sense; but it *is* a probability distribution in its own special sense. We use it to make statements of the kind: among the values of μ which are possible, only those in a certain range give rise to the observed $\bar{x}$ with probability α, and hence we will locate μ in that range.

20.5. In our present example the argument would proceed as follows. From equation (20.6) and the use of the normal integral, the probability that $\mu - \bar{x}$ does not exceed a certain h is ascertainable as a function of h; for instance,

$$P\left\{\mu - \bar{x} \leqslant \frac{2}{\sqrt{n}}\right\} = 0{\cdot}9775.$$

If we regard a probability as high as this as acceptable, we may say that $\mu \leqslant \bar{x} + 2/\sqrt{n}$.

This result is equivalent to that given by the theory of confidence intervals, for if we assert $\mu \leqslant \bar{x} + 2/\sqrt{n}$ we shall be right in the long run in 97·75 per cent. of the cases. This identity of result is found in most elementary cases where a single parameter is concerned, but is to be regarded as accidental. In the theory of confidence intervals it is fundamental (*a*) that the assertion as to the parameter lying in a given range should be true in an assigned proportion α of the cases, and (*b*) that no assumption need be made as to the prior distribution of the parameter, either in the frequency sense or in the fiducial sense. In fiducial theory it is not necessary that (*a*) should be true, but the fiducial distribution is a fundamental part of the inference.

20.6. There is a further distinction between the two theories. In that of confidence intervals it is possible to have two entirely different sets for the same parameter, and in fact part of that theory is devoted to finding " best " sets among the possible ones. In fiducial theory such a state of affairs must not be possible, for different limits would imply different fiducial distributions for the same parameter *on the same evidence*. This is avoided by confining fiducial distributions to those based on sufficient estimators, or more generally on a set of estimators which together avoid all loss of information. Since such estimators alone contain all the information relevant to the problem of estimation they alone can give the fiducial distributions accurately. It follows, of course, that where no sufficient

estimator—or estimator with complete set of ancillary estimators—can be found, the fiducial method is inapplicable.

20.7. Generally, let $F(\theta, t)$ be the distribution function of a sufficient estimator t for a parameter θ. Then for the frequency distribution of t we have

$$dF = \frac{\partial F(t, \theta)}{\partial t}\, dt. \qquad (20.7)$$

$F(t, \theta)$ is the probability that a random value of the estimator does not exceed a given value t. In accordance with the fiducial principle, this may be equated to the probability that for fixed t the value of θ will exceed t, so that for the fiducial distribution of θ we have

$$\begin{aligned} dF &= \frac{\partial}{\partial \theta}\{1 - F(t, \theta)\}\, d\theta \\ &= -\frac{\partial F(t, \theta)}{\partial \theta}\, d\theta. \qquad (20.8) \end{aligned}$$

This shows the general relation between the frequency-distribution of the estimator and the fiducial distribution of the parameter.

Example 20.1

If p is known, the estimator $\hat{\theta} = \frac{\bar{x}}{p}$ is sufficient for θ in samples from

$$dF = \frac{x^{p-1}\, e^{-x/\theta}}{\theta^p\, \Gamma(p)}\, dx, \qquad 0 \leqslant x \leqslant \infty$$

the distribution of $\hat{\theta}$ being, in fact,

$$dF = \left(\frac{np}{\theta}\right)^{np} \frac{\hat{\theta}^{np-1}}{\Gamma(np)} \exp\left(-\frac{np\hat{\theta}}{\theta}\right) d\hat{\theta}.$$

(Cf. Example 17.8.) We may write this in the form

$$dF = \left(\frac{np\hat{\theta}}{\theta}\right)^{np-1} \frac{\exp\left(-\dfrac{np\hat{\theta}}{\theta}\right)}{\Gamma(np)}\, d\left(\frac{np\hat{\theta}}{\theta}\right). \qquad (20.9)$$

It is then clear that, since

$$-\frac{\partial F}{\partial \theta} = -\frac{\partial F}{\partial t}\frac{\partial t}{\partial \theta},$$

the corresponding fiducial distribution of θ is

$$dF = \left(\frac{np\hat{\theta}}{\theta}\right)^{np-1} \frac{\exp\left(-\dfrac{np\hat{\theta}}{\theta}\right)}{\Gamma(np)}\, np\hat{\theta}\frac{d\theta}{\theta^2}, \qquad (20.10)$$

which may also be put in the form (20.9), provided that we interpret the differential element now as relating to θ and not to $\hat{\theta}$. It will be noticed that we have replaced $d\hat{\theta}$ by $\hat{\theta}\frac{d\theta}{\theta}$, not merely by $d\theta$.

From the fiducial distribution (20.10) we can find the probability that θ lies in a certain range dependent on the observed $\hat{\theta}$ and the chosen probability α. This is in fact the same range that we should obtain by applying confidence intervals to (20.9). Once again the results of the two methods are the same.

Fiducial Inference based on " Student's " Distribution

20.8. Consider now the estimation of the mean μ in samples from a normal population with unknown variance σ^2. The treatment of **20.2** is no longer of use, for it would result in a fiducial distribution of μ containing the unknown σ. We therefore " studentise " the problem by considering the distribution of

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s'} \qquad (20.11)$$

which is independent of σ, being in fact

$$dF \propto \frac{dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}, \qquad (20.12)$$

where $\nu = n - 1$. Here s'^2 is the unbiassed estimate of the sample variance

$$\frac{1}{n-1} \Sigma (x - \bar{x})^2.$$

The distribution of t may be written

$$dF \propto \frac{d\left\{\frac{\bar{x} - \mu}{s'}\sqrt{n}\right\}}{\left\{1 + \frac{(\bar{x} - \mu)^2 n}{s'^2 (n-1)}\right\}^{\frac{1}{2}n}}. \qquad (20.13)$$

The fiducial distribution is then

$$dF \propto \frac{d\mu}{\left\{1 + \frac{(\mu - \bar{x})^2 n}{s'^2 (n-1)}\right\}^{\frac{1}{2}n}}. \qquad (20.14)$$

In the usual way we can find two constants, for any given α, such that, from (20.14),

$$P\{\mu_0 \leqslant \mu \leqslant \mu_1\} = \alpha, \qquad (20.15)$$

the probability being based on (20.14) and therefore to be understood in the fiducial sense. Had we worked with (20.12) or (20.13) we should have found t_1, t_0 such that

$$P\{-t_1 \leqslant t \leqslant t_0\} = \alpha,$$

which is equivalent to

$$P\left\{\bar{x} - \frac{s' t_0}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + \frac{s' t_1}{\sqrt{n}}\right\} = \alpha. \qquad (20.16)$$

This may be interpreted in the sense of confidence intervals, i.e. that in asserting the inequality in (20.16) we should be right in a proportion α of the cases in the long run. (20.15) does not rest on this statement as to frequency, though the limits to which it leads are the same and the statement happens to be true.

20.9. The case we have just discussed raises a new point. Is it still true that the fiducial distribution is unique, and is it consistent with the distributions of μ and σ separately ? The distribution is based only on the sufficient estimators $\bar{x}$ and s' (which are jointly but not separately sufficient for μ and σ) and we should expect this to be so. But the matter requires investigation, for we are here using a fiducial distribution based on two estimators.

The simultaneous distribution of $\bar{x}$ and s' is

$$dF \propto \frac{1}{\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} d\bar{x} \left(\frac{s'}{\sigma}\right)^{n-2} \exp\left\{-\frac{(n-1)s'^2}{2\sigma^2}\right\} \frac{ds'}{\sigma}. \quad . \quad (20.17)$$

If we were considering fiducial limits for μ with *known* σ we should use the distribution

$$dF \propto \frac{1}{\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} d\bar{x}.$$

If we were considering fiducial limits for σ with known μ we should *not* use the other factor in (20.17),

$$dF \propto \left(\frac{s'}{\sigma}\right)^{n-2} \exp\left\{-\frac{(n-1)s'^2}{2\sigma^2}\right\} \frac{ds'}{\sigma}, \quad . \quad . \quad . \quad (20.18)$$

for in such circumstances s' is not sufficient for σ, the appropriate estimator being $\frac{1}{n}\Sigma(x-\mu)^2$. The question is, what form of fiducial distribution must hold for σ in order that the "Student" form (20.14) should hold for μ when σ is unknown?

Suppose the fiducial distribution is $f(s', \sigma)\, d\sigma$. We have then for the joint fiducial distribution of μ and σ,

$$dF \propto \frac{1}{\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} d\mu\, f(s', \sigma)\, d\sigma.$$

We have therefore to solve

$$\left\{\int_0^\infty \frac{1}{\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right\} f(s', \sigma)\, d\sigma\right\} d\mu = \frac{k\, d\mu}{\left\{1+\frac{(\mu-\bar{x})^2 n}{s'^2(n-1)}\right\}^{\frac{1}{2}n}} \quad . \quad (20.19)$$

where k is some constant. Putting $(\mu-\bar{x})^2 = \alpha$, $-\frac{n}{2\sigma^2} = \beta$, we have then to solve

$$\int_0^\infty e^{\alpha\beta} f\left(s', \sqrt{-\frac{n}{2\beta}}\right) \frac{d\beta}{\beta} = \frac{k}{\left\{1+\frac{n\alpha}{(n-1)s'^2}\right\}^{\frac{n}{2}}}.$$

Regarding α as the complex quantity it we see that $\frac{1}{\beta} f\left(s', \sqrt{-\frac{n}{2\beta}}\right)$ is the frequency function whose characteristic function is $1\Big/\left\{1+\frac{n\alpha}{(n-1)s'^2}\right\}^{\frac{n}{2}}$, which gives

$$\frac{1}{\beta} f\left(s', \sqrt{-\frac{n}{2\beta}}\right) \propto \beta^{\frac{1}{2}n-1} \exp\left\{\frac{(n-1)s'^2}{n}\beta\right\},$$

from which we find

$$f(s', \sigma) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{(n-1)s'^2}{2\sigma^2}\right\},$$

or, on evaluation of the constant,

$$f(s', \sigma)\, d\sigma = \frac{2}{\Gamma\left(\frac{n-1}{2}\right)} \left\{\frac{(n-1)s'^2}{2\sigma^2}\right\}^{\frac{1}{2}(n-1)} \exp\left\{-\frac{(n-1)s'^2}{2\sigma^2}\right\} \frac{d\sigma}{\sigma} \quad . \quad (20.20)$$

This, then, is the fiducial distribution which σ must obey. We should have arrived at

the same result had we taken (20.18) and transformed it to the fiducial form, as if it related to s' and σ only and the former were sufficient for the latter.

It appears, then, that in this case at least the fiducial method gives consistent results when two parameters are involved. The general problem of many parameters presents difficulties and has not been elucidated to any great extent.

The Logic of Fiducial Inference

20.10. The notion of fiducial probability was introduced by Fisher (1930) for the case of a single parameter. Regarding the estimate t as fixed, Fisher considers the distribution of values of θ for which t can be regarded as a representative estimate—representative, that is to say, in the sense that it could have arisen by random sampling from the population specified by θ. As pointed out above, this does not mean that we are regarding the true value of θ as a member of an existing population. Rather, we are considering the possible values of θ and attaching to each value a measure of our confidence in it, based on the probability that it could have given rise to the observed t.

If I interpret him correctly, Fisher would regard a fiducial distribution as a frequency-distribution. This implies that θ is regarded as a random variable. It appears to me, however, that it is not a random variable in the ordinary sense of the frequency theory of probability, in which values of θ either are or can be generated by an actual sampling process. We can never test whether the fiducial distribution holds in the frequency sense by drawing a number of values and comparing observation with theory. Nor, in calculating fiducial limits of the type $\theta = t + h(\alpha)$, do we imply that the proportion of cases for which $\theta \leqslant t + h$ is true will be α in the long run.

20.11. The reader has a choice of several attitudes towards the foundations of the fiducial argument: (*a*) he can accept the argument as involving a new postulate of inference; (*b*) he can regard it as sanctioned by the approach of the previous section; or (*c*) he can, so far as estimates based on a single parameter are concerned, console himself with the thought that the results of the process are the same as those given by the theory of confidence intervals.

20.12. Although Fisher is careful to emphasise the distinction between his own approach and that based on Bayes' postulate, it is interesting to note that the theory of inverse probability as modified by Jeffreys gives results which are in many cases identical with those of fiducial inference.

In the example of **20.2**, for instance, suppose that the prior distribution of μ is $f(\mu)\,d\mu$. Then for any given $\bar{x}$ the posterior probability of μ is

$$dF = f(\mu)\,d\mu \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\}. \qquad (20.21)$$

If the total probability is unity we have

$$\int_{-\infty}^{\infty} f(\mu) \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\} d\mu = 1. \qquad (20.22)$$

Clearly $f(\mu) = 1$ is a solution, and we may use characteristic functions to show that it is the only solution. In fact we have from (20.22), writing it for $n\bar{x}$—

$$\int_{-\infty}^{\infty} f(\mu) \exp\left(-\frac{n\mu^2}{2}\right) e^{it\mu}\,d\mu = \sqrt{\frac{2\pi}{n}} \exp\left(-\frac{t^2}{2n}\right).$$

The expression on the right is the characteristic function of $\exp\left(-\frac{n\mu^2}{2}\right)$, and hence

$$f(\mu)\exp\left(-\frac{n\mu^2}{2}\right) = \exp\left(-\frac{n\mu^2}{2}\right),$$

or $f(\mu) = 1$.

We have, then, for the posterior probability distribution of μ,

$$dF = \sqrt{\frac{n}{2\pi}}\exp\left\{-\frac{n}{2}(\mu - \bar{x})^2\right\}d\mu, \qquad . \quad . \quad . \quad (20.23)$$

which is the same as the fiducial distribution. The requirement that $f(\mu) = 1$ is equivalent to a prior distribution of μ, $dF = d\mu$, which is the form given by Bayes' postulate for a parameter which can extend to infinity in either direction.

Example 20.2

In Example 20.1, a similar argument leads to a prior distribution of θ,

$$dF \propto \frac{d\theta}{\theta}.$$

This is the form given by Jeffreys' modification of Bayes' postulate when a parameter can extend to infinity in only one direction.

It does not appear, however, that fiducial and inverse probability always give the same results. Consider the distribution of the correlation coefficient in normal samples **(14.14)**—

$$dF \propto (1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}\frac{d^{n-2}}{d(r\rho)^{n-2}}\left\{\frac{\cos^{-1}(-\rho r)}{\sqrt{(1-\rho^2 r^2)}}\right\}dr. \qquad . \quad (20.24)$$

The argument of the type we have just employed would require a prior distribution of ρ—

$$dF \propto \frac{d\rho}{(1-\rho^2)^{\frac{3}{2}}},$$

and the resulting posterior distribution (which is equivalent to that obtained by interchanging r and ρ in (20.24)) is not the same as we should get by using equation (20.8).

Behrens' Test

20.13. Suppose we have two samples of n_1 and n_2 members from normal populations with possibly unequal variances. The fiducial distributions of μ_1 and μ_2 are of the "Student" form (20.14). Writing

$$\mu_1 = \bar{x}_1 + s_1' u_1$$
$$\mu_2 = \bar{x}_2 + s_2' u_2$$

we have

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 + s_1' u_1 - s_2' u_2. \qquad . \quad . \quad . \quad . \quad (20.25)$$

If now

$$\varepsilon = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{(s_1'^2 + s_2'^2)}}, \qquad . \quad . \quad . \quad . \quad . \quad (20.26)$$

ε depends only on the known quantities $\bar{x}$ and s' and the difference of means $\mu_1 - \mu_2$. From the fiducial distributions of μ_1 and μ_2 we can find that of ε, and hence make fiducial statements of the type

$$\bar{x}_1 - \bar{x}_2 - \varepsilon_0\sqrt{(s_1'^2 + s_2'^2)} \leqslant \mu_1 - \mu_2 \leqslant \bar{x}_1 - \bar{x}_2 + \varepsilon_1\sqrt{(s_1'^2 + s_2'^2)}. \qquad . \quad (20.27)$$

20.14. The distribution of ε is not of a simple form. Putting $\tan\psi = \dfrac{s_2'}{s_1'}$ we see that

$$\varepsilon = \frac{\bar{x}_1 - \mu_1}{s_1'} \cos\psi - \frac{\bar{x}_2 - \mu_2}{s_2'} \sin\psi, \qquad . \quad . \quad . \quad (20.28)$$

so that ε is distributed fiducially as the weighted difference of two variables, each of which is distributed as " Student's " t. We have then to find the distribution of

$$\varepsilon = t_1 \cos\psi - t_2 \sin\psi$$

where the joint distribution of t_1 and t_2 is given by

$$dF \propto \frac{dt_1}{\left(1 + \dfrac{t_1^2}{n_1 - 1}\right)^{\frac{1}{2}n_1}} \frac{dt_2}{\left(1 + \dfrac{t_2^2}{n_2 - 1}\right)^{\frac{1}{2}n_2}} \qquad . \quad . \quad . \quad (20.29)$$

The distribution has been studied by Sukhatme (1938*b*) and in more detail by Fisher (1941*a*). Tables are given for various values of n_1, n_2 and the ratio $s_1'^2/s_2'^2$ (or the equivalent angle ψ) showing the values of ε corresponding to given probability levels. Some of the tables are included in the second (1943) edition of Fisher and Yates' *Statistical Tables for Agricultural, Biological and Medical Research.*

20.15. The joint distribution of $s_1'^2$ and $s_2'^2$ is

$$dF \propto s_1'^{n_1-3}\, s_2'^{n_2-3} \exp\left\{ -\tfrac{1}{2}(n_1 - 1)\frac{s_1'^2}{\sigma_1^2} - \tfrac{1}{2}(n_2 - 1)\frac{s_2'^2}{\sigma_2^2} \right\} ds_1'^2\, ds_2'^2.$$

Putting
$$p = \frac{s_1'^2}{s_2'^2} \quad \text{and} \quad u = \tfrac{1}{2}\left\{(n_1 - 1)\frac{s_1'^2}{\sigma_1^2} + (n_2 - 1)\frac{s_2'^2}{\sigma_2^2}\right\},$$

we find, on a little reduction,

$$dF \propto \frac{p^{\frac{1}{2}(n_1-3)}\, dp}{\left\{\dfrac{p\,(n_1 - 1)}{\sigma_1^2} + \dfrac{n_2 - 1}{\sigma_2^2}\right\}^{\frac{1}{2}(n_1+n_2-2)}} u^{\frac{1}{2}(n_1+n_2-4)} e^{-u}\, du. \qquad . \quad (20.30)$$

Thus u is distributed (independently of p) in the Type III form. Further, $(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)$ is distributed normally about zero mean with variance $\sigma_1^2 + \sigma_2^2$. Hence, if $\dfrac{\sigma_1^2}{\sigma_2^2} = \theta$, we find that the quotient

$$\frac{\{(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)\}^2 (n_1 + n_2 - 2)}{(\sigma_1^2 + \sigma_2^2)\left\{\dfrac{(n_1 - 1)\, s_1'^2}{\sigma_1^2} + \dfrac{(n_2 - 1)\, s_2'^2}{\sigma_2^2}\right\}} = \frac{\varepsilon^2 (1 + p)(n_1 + n_2 - 2)}{\left\{(n_2 - 1) + (n_1 - 1)\dfrac{p}{\theta}\right\}(1 + \theta)} \qquad (20.31)$$

is distributed as t^2 with $n_1 + n_2 - 2$ degrees of freedom. (Cf. Example 10.17, vol. I, p. 248, for the distribution of a normal variate divided by a Type III variate.)

Now if we knew θ we could find fiducial (or confidence) limits to ε, and hence to $\mu_1 - \mu_2$, in the usual way, for the distribution of ε would then be independent of unknown constants and ascertainable from " Student's " integral. Since, however, θ is not known, we require in turn the fiducial distribution of this quantity. Since

$$z = \tfrac{1}{2}\log\left(\frac{n_1 s_1'^2}{\sigma_1^2} \bigg/ \frac{n_2 s_2'^2}{\sigma_2^2}\right)$$

is distributed in Fisher's form (cf. Example 10.18, vol. I, p. 249), the required fiducial

form for θ can be obtained from that of z, which incidentally is equivalent to that of p in (20.30). If we express (20.31) as the joint fiducial distribution of ε and θ and integrate out for θ, we shall be left with an equivalent form to that derived from (20.29).

20.16. It also follows from the above that the inequality (20.27) is *not* satisfied in proportion α of the cases independently of θ, so that the limits to $\mu_1 - \mu_2$ are not confidence limits, although they are fiducial limits. It will, in fact, be evident enough from (20.31) that if we determine t_0 and t_1 so that the integral of "Student's" form between those limits is α, then the corresponding limits for ε, say ε_0 and ε_1, are dependent on the variance ratio $\theta = \sigma_1^2/\sigma_2^2$. This is fairly evident on general grounds, and the point has been put beyond doubt by both Fisher (1937*b*) and Neyman (1941*a*), who have worked out particular cases of difference.

The fiducial distribution of ε (which is an extension by Fisher of a result given by Behrens as early as 1929) thus provides a crucial point of difference between the theory of fiducial inference and that of confidence intervals.

20.17. In conclusion, we will indicate the viewpoint of Jeffreys towards the type of problem dealt with by "Student's" distribution for limits to the mean and Behrens' distribution for limits to the difference of two means.

If H denotes the general data, we have for the "Student" distribution—

$$P\{dt \mid \mu, \sigma, H\} = \frac{k\,dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}. \quad . \quad . \quad . \quad . \quad (20.32)$$

The expression on the left states the probability that t will lie in a given range dt on the assumption that H is true, the parent mean being μ and the parent variance σ^2. Since μ and σ do not appear on the right they are irrelevant and may be suppressed, and hence

$$P\{dt \mid H\} = \frac{k\,dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}. \quad . \quad . \quad . \quad . \quad (20.33)$$

Suppose now that we *assume* that

$$P\{dt \mid \bar{x}, s, H\} = f(t)\,dt. \quad . \quad . \quad . \quad . \quad . \quad (20.34)$$

Then, as before, $\bar{x}$ and s may be suppressed and we have

$$P\{dt \mid H\} = f(t)\,dt, \quad . \quad . \quad . \quad . \quad . \quad . \quad (20.35)$$

and hence, by comparison with (20.33),

$$P\{dt \mid \bar{x}, s, H\} = \frac{k\,dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}. \quad . \quad . \quad . \quad . \quad (20.36)$$

We can then proceed to find limits to t, given $\bar{x}$ and s, in the usual way. Jeffreys emphasises, however, that this depends on a new postulate expressed by (20.34) which, though natural, is not trivial. It amounts to an assumption that if we are comparing different distributions, samples from which give different $\bar{x}$'s and s's, the scale of the distribution of μ must be taken proportional to s and its mean displaced by the difference of sample means.

20.18. In a similar way it will be found that to arrive at the Behrens distribution it is necessary to postulate that

$$P\{dt_1, dt_2 \mid \bar{x}_1, \bar{x}_2, s_1', s_2', H\} = f_1(t_1) f_2(t_2)\, dt_1\, dt_2 \quad . \quad . \quad . \quad (20.37)$$

Jeffreys' derivation of the Behrens' form from Bayes' theorem would be as follows :—

The prior probability of $d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2 \mid H$ is

$$P\{d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2 \mid H\} \propto \frac{d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2}{\sigma_1\, \sigma_2}.$$

The likelihood (denoting the data by D) is

$$P\{D \mid \mu_1, \mu_2, \sigma_1, \sigma_2, H\} \propto \frac{1}{\sigma_1^{n_1}\, \sigma_2^{n_2}} \exp\left[-\frac{n_1}{2\sigma_1^2}\{(\mu_1 - \bar{x}_1)^2 + s_1^2\} - \frac{n_2}{2\sigma_2^2}\{(\mu_2 - \bar{x}_2)^2 + s_2^2\}\right].$$

Hence, by Bayes' theorem

$$P\{d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2 \mid DH\} = \frac{1}{\sigma_1^{n_1+1}\, \sigma_2^{n_2+1}} \exp\left[-\frac{n_1}{2\sigma_1^2}\{(\mu_1 - \bar{x}_1)^2 + s_1^2\}\right.$$
$$\left. - \frac{n_2}{2\sigma_2^2}\{(\mu_2 - \bar{x}_2)^2 + s_2^2\}\right] d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2.$$

Integrating out the values of σ_1 and σ_2, we find for the posterior distribution of μ_1 and μ_2 a form which is easily reducible to (20.29).

20.19. To sum up : so far as concerns problems of estimation the Behrens test is accurate both in fiducial theory and in the theory of probability propounded by Jeffreys. But the test does not hold in the theory of confidence intervals. In fact the latter fails to provide an exact solution to the problem, though we shall see below (**21.28**) that approximations are possible. Fisher has criticised confidence intervals on the ground that they do not give an answer to what is admittedly an important question ; but it appears possible to maintain consistently that some questions may not have an answer.

NOTES AND REFERENCES

For the general theory of fiducial inference see Fisher (1930*a*, 1933, 1935*a*, *b*, 1936*c*, 1941*a*). The difficulties of reconciling Behrens' test with confidence-interval theory were noticed by Bartlett (1936*a*) and led to some controversy, for which see Fisher (1937*b*, 1939*a*, 1940*c*), Bartlett (1939*a*), Yates (1939*f*), and Neyman (1941*a*). For Jeffreys' views see his papers of 1937*b*, 1938*c*, 1939*d* and 1940.

For the practical application of Behrens' distribution see Sukhatme (1938*b*) and Fisher (1941*a*). Behrens himself stated his results explicitly only for the case of equality of sample number, $n_1 = n_2$, the extension being given by Fisher (1935*b*).

EXERCISES

20.1. If $\bar{x}$ is the mean of a sample of n values from

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx,$$

s'^2 is equal to $\dfrac{1}{n-1}\Sigma(x - \bar{x})^2$, and x is a further independent sample value, show that

$$t = \frac{x - \bar{x}}{s'}\sqrt{\frac{n}{n+1}}$$

is distributed in "Student's" form with $\nu = n - 1$. Hence show that fiducial limits for x are

$$\bar{x} \pm s' t_1 \sqrt{\frac{n+1}{n}},$$

where t_1 is chosen so that the integral of "Student's" form between $-t_1$ and t_1 is an assigned probability α.

(Fisher, 1935*b*. This gives an estimate of the next value when n values have already been chosen, and extends the idea of fiducial limits from parameters to variates dependent on them.)

20.2. Show similarly that if a sample of n_1 values gives mean $\bar{x}_1$ and estimated variance $s_1'^2$, the fiducial distribution of mean $\bar{x}_2$ and estimated variance $s_2'^2$ in a second sample of n_2 is

$$dF \propto \frac{s_1'^{\,n_1-1} s_2'^{\,n_2-2} \, d\bar{x}_2 \, ds_2'}{\left\{(n_1 - 1) s_1'^2 + (n_2 - 1) s_2'^2 + (\bar{x}_1 - \bar{x}_2)^2 \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}\right\}^{\frac{1}{2}(n_1+n_2-1)}}.$$

Hence, allowing n_2 to tend to infinity, derive the simultaneous fiducial distribution of μ and σ.

(Fisher, 1935*b*.)

CHAPTER 21

SOME COMMON TESTS OF SIGNIFICANCE

Tests of Significance

21.1. We now pass from the problem of estimation to that of significance. The two are closely allied and in practical problems they both arise together as a rule; but it is useful to preserve a distinction between them. In estimation we try to find, with greater or less accuracy, the value of some parameter in a population which is known to be (or assumed to be) dependent on that parameter. In tests of significance we are given some value of a parameter beforehand and wish to decide whether it is acceptable in the light of the evidence. This is the distinction in its simplest terms, but of course the associated problems become increasingly complex when several parameters are concerned.

21.2. From one point of view the problem of significance is logically anterior to that of estimation. Suppose we have records of the yields of two varieties of wheat grown under similar conditions, and are interested in a comparison of the average yields of the two. Our first question is whether the observed mean yields indicate any difference between the varieties—a matter of significance. Not until significant differences are established does our interest turn to the *magnitude* of the difference—a matter of estimation. Again, if we have a set of records of only one variety, our primary problem may be to decide whether they are consonant with the hypothesis of normality in the parent population, whatever its mean and variance; and only when this point has been settled affirmatively do we proceed to estimate those parameters.

Nevertheless, we have lost very little by taking the problem of estimation first. In some practical problems the question of significance is already decided, and in many others we use estimates of parameters to test the significance of the latter, in which case estimation and significance become different aspects of the same statistical fact.

21.3. We shall consider the general theory of testing statistical hypotheses in Chapters 26 and 27. That theory is, however, rather abstract, and we anticipate it to some extent in this chapter by giving an account of the principal tests in current use, without for the moment going too deeply into their rationale. It will be seen later that there are sometimes many significance tests which can be applied to the same problem, and that it is possible to lay down criteria for deciding which, if any, are the " best ". This aspect of the subject will not concern us for the present. We shall not discuss whether the tests we describe are the best possible (though some of them, in fact, are so) but shall merely present them as useful and convenient, albeit perhaps not unique, solutions of our problems.

21.4. Developments in statistical theory in the last two decades have resulted in a great many tests of significance appropriate to special problems. It is not easy to classify them and quite impossible to deal extensively with them all. We shall consider them under the following heads:—

(*a*) *Tests of the significance of a specified parameter value.*—The typical hypothesis here is that a parameter in a population of known form has a specified value (usually zero). We wish to know whether the evidence provided by the sample supports the hypothesis or not.

(*b*) *Tests of goodness of fit.*—The hypothesis is that the population is of a certain kind which is either fully specified beforehand or can be "estimated" with the help of the data. We wish to know whether the sample values fit this population in the sense that they could have arisen from it by random sampling to any acceptable degree of probability. This hypothesis is more general than that of (*a*) since it concerns the whole distribution function and not merely one of its parameters.

(*c*) *Tests of homogeneity.*—The hypothesis here concerns two or more populations, each providing a contribution to the sample. We wish to test whether the populations have certain parameters in common, or in the extreme case, whether they are identical. This case can be regarded as an elaboration of (*a*) where several parameters are simultaneously tested. In the particular case when only two populations are concerned we may sometimes reduce it directly to type (*a*) by considering differences; e.g. if we are making a comparison of parent means the hypothesis might be that the single difference of means is zero.

In addition we shall also consider two sets of tests of rather a different kind:—

(*d*) *Tests of order of occurrence.*—The hypothesis here is that the sample members occurred in random order, and we wish to ascertain whether the observed order indicates any systematic effects, as, for instance, whether there are any cyclical effects in time-series. The test here is of the sampling process rather than of parameters of the parent population.

(*e*) *Conditional tests.*—The hypothesis may be any one of the above types, but we restrict the inference to a sub-population for which certain qualities are determined by the observed sample values. For instance, we may use the distribution of the sample variance s^2 for which the mean $\bar{x}$ is equal to the observed value. In short the variation of sample values is *conditioned.* Type (*d*) may from some points of view be regarded as a particular case of this type.

It is not intended to convey that the above five categories are mutually exclusive. A test of type (*a*) may, for example, be conditional or non-conditional. The classification will, however, provide some sort of articulation for a rather long chapter and serve to explain our sequence of treatment.

Standard Errors

21.5. For large samples the test of significance of a parameter can usually be carried out by standard errors. We find an estimator t of the parameter θ and consider whether the given value of θ falls in the range $t_1 \pm k\sqrt{\text{var } t}$, where t_1 is the value of t for the observed sample and k is a constant chosen at will according to a probability α. If so we may accept the value of θ, at least so far as this test is concerned; if not, we reject it.

If the variance of t does not depend on unknown quantities such as other parameters, this type of inference is justifiable as an application of the theory of confidence intervals. In accepting θ when it falls in the range $t_1 \pm k\sqrt{\text{var } t}$, we shall be right in proportion α of the cases in the long run. As a refinement we may, of course, use non-central intervals and locate θ in an asymmetrical range $t_1 - k_0\sqrt{\text{var } t}$ to $t_1 + k_1\sqrt{\text{var } t}$. The test of significance is equivalent to the estimation of the true value of θ; and it will clearly be better if the range of estimation is narrower, for then we reject more wrong values of θ.

21.6. If the variance of the estimator t depends on unknown parameters $\theta_2 \ldots \theta_p$ we can usually substitute estimates of those parameters obtained from the sample itself,

provided that the sample is large. For example, we have for normal samples

$$P\left(\mu \leqslant \bar{x} + \frac{2\sigma}{\sqrt{n}}\right) = 0{\cdot}97725.$$

The sample standard deviation s will differ from σ by a quantity of order $1/\sqrt{n}$, so that to that order

$$P\left\{\mu \leqslant \bar{x} + \frac{2s}{\sqrt{n}}\right\} = 0{\cdot}97725.$$

The approximation breaks down for small samples, and more accurate methods are required.

21.7. The use of standard errors in testing significance has been illustrated in previous chapters, and we need not enlarge on the process further. We may, however, remark two things :—

(*a*) That if the distribution of an estimator t tends to normality for large samples irrespective of the parent form (as, for instance, is the case with the mean and other moments under very general conditions), it is not necessary that the hypothesis should specify the parent form. In short, our test of significance is independent of the parent, a valuable generality which rarely obtains for small samples.

(*b*) That we have justified the logic of reasoning involving the use of standard errors by the theory of confidence intervals (and a similar justification can be given in terms of fiducial intervals if we use an efficient estimator for which the loss of information tends to zero relative to the total information in large samples). This appears to be the most satisfactory basis for the use of standard errors. The usual intuitive basis advanced (necessarily) in introductory textbooks is not easy to defend. For instance, it is customary to reject a value of θ if it gives to an observed t_1 *or greater value* a small probability ; and there is no obvious reason why we should base our inference on the improbability of greater values of t_1, namely on the improbability of something which has not occurred (see **21.55** below). Our present approach shows that in fact the use of standard errors can be justified logically without invoking a new principle of inference.

Significance of the Mean in Normal Samples

21.8. Suppose we have a sample from a parent population which is known to be normal, but of whose mean and variance we are ignorant. We wish to test the significance of a given value μ_0 of the mean, that is to say, we wish to consider whether the observations could, to any acceptable probability, have been derived from a population with mean μ_0, whatever the variance may be.

We calculate the statistic

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{\nu}, \qquad (21.1)$$

all the quantities in which are given. We know that the distribution of t is

$$dF = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\pi\nu)}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}, \qquad (21.2)$$

and hence can find the probability that our calculated value of t is attained or exceeded. If this is small we reject μ_0 ; if not, we accept it. What values are regarded as " small "

for this purpose is a matter of convention, but the most frequently used values are 0·05, 0·01 and 0·001.

From the work of the previous two chapters it will be evident that this type of inference is the confidence- or fiducial-interval approach in a slightly different form. Given α we can find $-t_1$ and t_0 such that the integral of dF in (21.2) between those limits is α. This gives us confidence or fiducial limits to μ of the type $\bar{x} - \frac{t_0 s}{\sqrt{\nu}}$ and $\bar{x} + \frac{t_1 s}{\sqrt{\nu}}$; and if μ_0 lies in this range we accept it. In particular cases we may have $t_0 = t_1$, in which cases the intervals are central and our probability α is the chance of t being attained or exceeded *in absolute value*; or $t_0 = +\infty$, in which case α is the chance that $-t_1$ will be attained or exceeded, and no lower limit to μ_0 is imposed.

Example 21.1

The weights of fifteen bags of sugar taken from a filling machine are found to be, in ounces, 16·1, 15·8, 15·8, 15·9, 16·1, 16·2, 16·0, 15·9, 16·0, 15·7, 15·7, 15·8, 16·0, 16·0, 15·8. Each bag should be 16 ounces, but some deviation is inevitable. One of the manufacturer's problems, of course, is to keep this deviation to a minimum, but that is not the point we now consider. Our question is: if the machine is supposed to be giving weights of 16 ounces on the average, does the sample suggest that it is failing in its purpose?

The hypothesis is that the parent mean is 16 ounces and the deviations from this mean are, in order of magnitude, $-$ 0·3 (twice), $-$ 0·2 (four times), $-$ 0·1 (twice), 0·0 (four times), 0·1 (twice), 0·2 (once). The sample mean is thus $-$ 0·08 and to that extent the average of the sample is slightly underweight. Is this a significant effect?

It will be found that $s^2 = 0{\cdot}0216$ so that

$$t = -\frac{0{\cdot}08}{\sqrt{0{\cdot}0216}}\sqrt{14} = -2{\cdot}04, \qquad \nu = 14.$$

From Appendix Table 3 (vol. I, p. 440) we find that for $\nu = 14$ the probability of a deviation greater in absolute magnitude than 2·04 is about 2 (1 $-$ 0·969) = 0·062. This is small, but whether we regard it as significant or not depends on the probabilities we are prepared to consider as defining significance. The usual values are 0·05 and 0·01, and with such criteria we should not take the observed value as significant, though it arouses suspicions.

We have here used central intervals, which are usual for the t-test of significance of the mean; but it is easy to imagine circumstances in this particular case for which non-central intervals might be required. For instance, if the machine was at fault and had a true mean filling weight of more than 16 ounces the manufacturer would be giving sugar away for nothing. This might be serious, but probably not so serious as if the machine was erring in the other direction, which would render him liable to prosecution for selling short weight. Suppose he assessed the latter risk as nine times as serious as the former and was working to a probability level of 0·05. Then he would require the probability of a *negative* value of t greater than the significance value to be 0·955 (= 1 $-$ 0·045) but could allow that of a positive value less than the significance value to be 0·995 (= 1 $-$ 0·005). From Appendix Table 3 we see that this corresponds to deviations of approximately $-$ 1·8 and + 3·0. Our observed value is outside this range and is thus significant. Small as the average shortage is, it would be prudent to overhaul the machine and to make sure that it is giving fair weight on the average.

We may note further that if the sample had occurred in the order

15·7, 15·7, 15·8, 15·8, 15·8, 15·8, 15·9, 15·9, 16·0, 16·0, 16·0, 16·0, 16·1, 16·1, 16·2

we should almost certainly have concluded that there was something wrong with the machine, for the weights are steadily rising. The t-test would give the same result for this sample as for the first, since it does not depend on the order of occurrence of the members. Where, therefore, the appearance of individual sample members is ordered in time, the t-test alone may fail to reveal significant effects due to the changing of the population between drawings. Our data are still such as could have arisen at a single drawing of fifteen members from a population with mean equal to 16 ounces; but the data throw doubt on the point whether we are really asking the right question in assuming that they all came from the same population. We consider the point again below (**21.41**).

Before leaving this example, we may note another possible test, cruder than the t-test but sometimes useful. If the parent mean were really zero, positive and negative deviations should occur equally frequently in the long run. In our present case there are 8 negative deviations, 3 positive ones and 4 zero. If we allot, conventionally, two of the last to each group we have 10 negative and 5 positive deviations. The expected number is $7\frac{1}{2}$, so that the deviation is $2\frac{1}{2}$, with a standard error of $\sqrt{(15 \times \frac{1}{2} \times \frac{1}{2})} = 1{\cdot}94$. The observed deviation is very little in excess of this, so we conclude that the preponderance of negative signs in the sample is not significant of a negative mean in the population. More exactly, we find that the occurrence of 5 or fewer positive deviations is the sum of the first six terms in the binomial $(\frac{1}{2} + \frac{1}{2})^{15}$, namely $0{\cdot}151$, leading to the same conclusion. The test is a very rough one since it pays no attention to the magnitude of the deviations; but it has the advantage of applying to any symmetrical form of parent population for finite samples.

Properties of the t-Distribution

21.9. "Student's" distribution has numerous applications in the testing of significance apart from the one just considered, and we proceed to study its properties.

The form (21.2) is a Pearson Type VII and may be transformed to the Beta-distribution (Type I) by the substitution $\xi = 1\Big/\left(1 + \frac{t^2}{\nu}\right)$. The distribution function of t may thus be obtained direct from the B-function. For instance, we have

$$F(t) = \int_{-\infty}^{t} dF = \tfrac{1}{2} + \int_0^t \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu\pi)}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{dt}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

whence

$$\begin{aligned} 2F - 1 &= \frac{2}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_0^t \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \frac{dt}{\sqrt{\nu}} \\ &= \frac{1}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{\xi}^{1} \xi^{\frac{\nu}{2}-1} (1 - \xi)^{-\frac{1}{2}}\, d\xi \\ &= 1 - I_{\xi}\left(\frac{\nu}{2}, \frac{1}{2}\right), \end{aligned}$$

whence

$$F = 1 - \tfrac{1}{2} I_{\xi}\left(\frac{\nu}{2}, \frac{1}{2}\right). \qquad (21.3)$$

The values of the argument for which I has the values 0·50, 0·25, 0·10, 0·05, 0·025, 0·01, 0·005 and $\nu = 1\ (1)\ 30, 40, 60, 120, \infty$, have been tabled to five significant figures by C. M. Thompson and others (1941a) and can hence be used to derive the values of t corresponding to those probability levels.

21.10. Except for special purposes, however, the use of the B-function is unnecessary, since the distribution function of t itself and tables based thereon are available.

We have

$$-\log\left(1+\frac{t^2}{\nu}\right) = -\frac{t^2}{\nu}+\frac{t^4}{2\nu^2}-\ldots+\frac{(-t^2)^j}{j\nu^j}+\ldots$$

and hence

$$-\frac{\nu+1}{2}\log\left(1+\frac{t^2}{\nu}\right) = -\tfrac{1}{2}t^2+\ldots+\frac{j(-t^2)^{j+1}+(j+1)(-t^2)^j}{2j(j+1)\nu^j}+\ldots \qquad (21.4)$$

Further, from the expansion for $\log \Gamma(1+x)$ we find

$$\log\left\{\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{2}{\nu}}\right\} = -\frac{1}{4\nu}+\frac{1}{24\nu^3}-\frac{1}{20\nu^5}\ldots \qquad (21.5)$$

Now as ν tends to infinity, t tends to the normal form with zero mean and unit variance. Writing

$$y = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}t^2},$$

we find for the logarithm of the ordinate of (21.2), in descending powers of ν,

$$\log y + \frac{1}{4\nu}(t^4-2t^2-1) - \frac{1}{12\nu^2}(2t^6-3t^4) + \frac{1}{24\nu^3}(3t^8-4t^6+1)$$
$$-\frac{1}{40\nu^4}(4t^{10}-5t^8)+\frac{1}{60\nu^5}(5t^{12}-6t^{10}-3)-\ldots \qquad (21.6)$$

Taking the exponential and integrating from t to ∞, we find

$$1-F = y\left\{\frac{1}{4\nu}t(t^2+1)+\frac{1}{96\nu^2}(3t^6-7t^4-5t^2-3)t+\frac{1}{384\nu^3}(t^{10}-11t^8\right.$$
$$+14t^6+6t^4-3t^2-15)t+\frac{1}{92160\nu^4}(15t^{14}-375t^{12}+2225t^{10}-2141t^8$$
$$\left.-939t^6-213t^4-915t^2+945)t+\ldots\right\} \qquad (21.7)$$

This is the expression, due to Fisher, which was used by "Student" himself in calculating the distribution function of t given in Appendix Table 3, Vol. 1. For values of $\nu \geqslant 18$ the first four terms of (21.7) give F to an accuracy of about 0·000,005.

21.11. Tables are also available in the "inverse" form, that is to say, giving values of t corresponding to specified values of ν and F. Such tables may be derived by interpolation from the "Student" tables or by the normalisation method of **6.32.** In work involving tests of significance this type of table is perhaps the most convenient, since it

enables one to decide without calculation (other than interpolation for values of the argument not covered by the tables) whether particular values are significant for chosen probability α. The complement of the probability α is spoken of as a level of significance and expressed either as a number between 0 and 1 or as a percentage. Similarly the corresponding values of t are called significance points, and we may speak, for example, of the 5 per cent. value of t, meaning that value for which F is 0·95.

Fisher and Yates (1938*a*) give the values of t for $\nu = 1\ (1)\ 30,\ 40,\ 60,\ 120$ and ∞ and $2\,(1 - F) = 0{\cdot}9\ (0{\cdot}1)\ 0{\cdot}1,\ 0{\cdot}05,\ 0{\cdot}02,\ 0{\cdot}01,\ 0{\cdot}001$. These tables, it should be remembered, give the significance points corresponding to *twice* $1 - F$, that is to say the values of t such that the proportion of the distribution outside the range $\pm\, t$ is $1 - F$.

21.12. The number ν is usually called the number of *degrees of freedom* of t. This is an expression which occurs in other connections, and a few words of explanation are desirable.

It has been seen that the variance of a normal sample is distributed like the sum of $(n - 1)$ squares of independent variates (compare Example 10.5, vol. I, p. 238) and generally, that if there are k linear relations connecting the original variates, the sum of squares of the originals is distributed as the sum of $n - k$ independent normal variates of equal variance. Each linear relation reduces the freedom of the variation, as it were, by unity. It is thus natural to speak of the number of degrees of freedom, ν, of a function such as χ^2, meaning thereby that it is distributed as the sum of squares of ν independent normal variates with equal variance. The expression only has this natural meaning when normal variation is concerned.

It so happens that the quantity t depends on a parameter ν which is convenient for tabulating its distribution function and is also the number of degrees of freedom of the statistic s^2 entering into the denominator of t. ν may thus, by an extension of the term, be called the number of degrees of freedom of t, but this usage does not imply that t is distributed as the sum of squares of normal variates.

Distribution of t in Non-normal Case

21.13. Part of the price we have to pay for the precision of the t-test in small samples is the assumption of normality in the parent. If the population is not normal we may still, of course, consider the distribution of " Student's " ratio, which will remain independent of the scale parameter ; but complications appear because the parameters which express the deviation from normality will, in general, appear in the sampling distribution. Furthermore, the distributions of $\bar{x}$ and s are no longer independent.

Let us in the first instance prove the last assertion which is due to Geary (1936*b*), in the form : If the mean and variance in samples from a population are independent and the population has finite cumulants, it must be normal.

From **11.13** we have

$$\kappa\,(21^r) = \frac{\kappa_{r+2}}{n^r}, \qquad r > 0.$$

If mean and variance are independent, $\kappa\,(21^r) = 0$ and hence $\kappa_{r+2} = 0$ for $r > 0$. Thus the population must be normal. It is rather remarkable that we have not had to use relations of the type $\kappa\,(2^s\,1^r) = 0$, $s > 1$ in arriving at this result and that we need only assume independence for one size of sample.

21.14. In the notation of Chapter 11 we write

$$t = \frac{\bar{x}\sqrt{\nu}}{s} = \frac{k_1\sqrt{\nu}}{\sqrt{\kappa_2}\left(1 + \dfrac{k_2 - \kappa_2}{\kappa_2}\right)^{\frac{1}{2}}},$$

and expand in terms of powers of $\dfrac{k_2 - \kappa_2}{\kappa_2}$. The method follows that of **11.23** and we find for the moments of t about the parent mean, assumed zero, to order ν^{-2}

$$\left.\begin{aligned}
\mu_1' &= -\frac{1}{\sqrt{\nu}}\left\{\tfrac{1}{2}\lambda_3 + \frac{3}{16\nu}(2\lambda_3 - 2\lambda_5 + 5\lambda_3\lambda_4\right\}\\
\mu_2' &= 1 + \frac{2}{\nu}(1 + \lambda_3^2) + \frac{2}{\nu^2}(3 - \lambda_4 - 3\lambda_3\lambda_5 + 6\lambda_3^2\lambda_4)\\
\mu_3' &= -\frac{1}{\sqrt{\nu}}\left\{\frac{7}{2}\lambda_3 + \frac{1}{16\nu}(210\lambda_3 - 66\lambda_5 + 105\lambda_3\lambda_4 + 210\lambda_3^3)\right\}\\
\mu_4' &= 3 + \frac{2}{\nu}(9 - \lambda_4 + 14\lambda_3^2) + \frac{1}{\nu^2}(102 - 30\lambda_4 + 24\lambda_5\\
&\qquad + 120\lambda_3^2 + 4\lambda_6 - 132\lambda_3\lambda_5 - 6\lambda_4^2 + 168\lambda_3^2\lambda_4 + 120\lambda_3^4)
\end{aligned}\right\} \quad . \quad (21.8)$$

where $\qquad \lambda_r = \dfrac{\kappa_r}{\kappa_2^{\frac{1}{2}r}}.$

If the parent form is symmetrical, cumulants of odd order vanish and we have, to order ν^{-2} and first order terms in the λ's—

$$\left.\begin{aligned}
\mu_1' &= \mu_3 = 0\\
\mu_2' &= 1 + \frac{2}{\nu} + \frac{6}{\nu^2} - \frac{2\lambda_4}{\nu^2} \ldots = \frac{\nu - 1}{\nu - 3} - \frac{2\lambda_4}{\nu^2}\\
\mu_4 &= 3 + \frac{18}{\nu} + \frac{102}{\nu^2} - \frac{2\lambda_4}{\nu} - \frac{30\lambda_4}{\nu^2} + \ldots = \frac{3(\nu - 1)^2}{(\nu - 3)(\nu - 5)} - \frac{2\lambda_4}{\nu} - \frac{30\lambda_4}{\nu^2}
\end{aligned}\right\} \quad (21.9)$$

Except for the term in λ_4 these are the values of the moments of t in "Student's" distribution, and it follows that for symmetrical parents which are not excessively lepto- or platykurtic we should not expect the t-test to be invalidated. If the parent is skew the situation may be different.

21.15. The general skew case has been considered by E. S. Pearson and Adyanthaya (1928, 1929) from the experimental viewpoint and by Bartlett (1935*a*) and Geary (1936*b*) from the theoretical viewpoint. Various writers have derived exact distributions of t in non-normal samples, but the sample numbers are, as a rule, trivially small and the results of little practical value. Geary considers the population expressed by the first two terms of the Gram-Charlier series—

$$dF = \frac{1}{\sqrt{2\pi}}\left\{1 - \frac{\kappa_3}{6}(3x - x^2)\right\} e^{-\frac{1}{2}x^2}\,dx \quad . \quad . \quad . \quad (21.10)$$

and assumes that powers of κ_3 above the first may be neglected. He finds (cf. Exercise **21.1**) that the frequency function of t in this population is equal to the "Student" form plus a corrective factor

$$\frac{\kappa_3}{6\nu}\,\frac{1}{\sqrt{\{2\pi(\nu + 1)\}}}\{3\nu - t^2(2\nu + 1)\}\,\frac{t\,dt}{\left(1 + \dfrac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+4)}}. \quad . \quad . \quad (21.11)$$

The integral of this factor from $-\infty$ to $-t$ is

$$\frac{\kappa_3}{6}\sqrt{\left(\frac{1}{2(\nu+1)\pi}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+2)}\left(1+\frac{2\nu+1}{\nu}t^2\right), \quad . \quad . \quad (21.12)$$

giving the correction to be applied. (Geary gives a table for some representative values.) This, of course, depends on κ_3, but even where exact knowledge of the skewness is not available we may sometimes safeguard against error by considering the correction for plausible values of κ_3.

Other Uses of the t-distribution

21.16. The usefulness of "Student's" t derives from the fact that it is independent of the scale parameter, and the simplicity of its distribution from the fact that it is the ratio of two *independent* variates, the numerator distributed normally and the denominator distributed in the Type III form. We shall see below (**21.26**) that these properties can be used to test the difference of two means in normal populations with equal variance, and in Chapter 22 we shall encounter a test of regression coefficients which is based on the same properties.

We have also noted that "Student's" form can be used to test the significance of the product–moment correlation (**14.15**) and the Spearman rank correlation ρ (**16.18**). These facts are, however, in a sense accidental. They do not derive from the expression of the parameters concerned as the ratio of a normal to a Type III variate, but from the simpler fact that the distributions are of the Type II form (symmetrical with finite range) and hence can be transformed to the "Student" distribution, which is of Type VII. Symmetrical distributions of finite range can often be represented very approximately by a transformation to the "Student" form, especially if they tend to normality.

Test of a Variance in Normal Samples

21.17. The distribution of the sample variance s^2 in normal samples is

$$dF=\frac{(\frac{1}{2}n)^{\frac{1}{2}(n-1)}}{\Gamma\left(\frac{n-1}{2}\right)}\left(\frac{s^2}{\sigma^2}\right)^{\frac{1}{2}(n-3)}\exp\left(-\frac{ns^2}{2\sigma^2}\right)d\left(\frac{s^2}{\sigma^2}\right) \qquad 0\leqslant s\leqslant\infty. \quad . \quad (21.13)$$

Thus, given for consideration a value of σ^2 and an observed s^2, we can find the probability that s^2/σ^2 is attained or exceeded and accept or reject σ^2 in the usual way. The distribution function of (21.13) may be expressed as an incomplete Γ-function, or more conveniently for statistical purposes in terms of $\chi^2\,(=ns^2/\sigma^2)$ with $\nu=n-1$.

Example 21.2

In Example 21.1 we found $s^2=0\cdot0216$, $\nu=14$. Could the data have arisen by chance from a population in which the true variance is 0·01 ?

We have $\chi^2=\dfrac{ns^2}{\sigma^2}=32\cdot4$, $\nu=14$. From the diagram on p. 446 of vol. I we see that the probability of such a value or greater is between 0·01 and 0·001, a very improbable result ; and hence we reject $\sigma^2=0\cdot01$ as a value of the parent variance.

Once again this type of inference can be justified by the theory of confidence intervals since the probability

$$P\left\{\frac{ns^2}{\sigma^2}\geqslant 32\cdot4\right\}<0\cdot01$$

is equivalent to

$$P\left\{\sigma^2 \leqslant \frac{ns^2}{32{\cdot}4}\right\} < 0{\cdot}01.$$

In asserting that σ^2 was less than $ns^2/32{\cdot}4$ (in our present case 0·01) we should be wrong more than 99 times in 100 on the average.

There is a point of interest to note here. In Example 21.1 we considered a hypothesis as to the mean μ, and in the present example a hypothesis as to the variance σ^2. Had we considered the two together, that is to say the compound hypothesis that $\mu = 16$ and $\sigma^2 = 0{\cdot}01$, we should have been in difficulties in justifying our procedure by reference to confidence or fiducial intervals, since we could no longer assert that our conclusions were right in an assigned proportion of cases. We have avoided this complication by considering separately the hypotheses (*a*) that $\mu = 16$ *whatever the variance*, and (*b*) that $\sigma^2 = 0{\cdot}01$ *whatever the mean*. This resource is not as a rule open to us where non-normal variation is concerned.

Tests of Normality

21.18. In large samples we can group the data into ranges and compare the actual frequencies with those to be expected on the hypothesis of parent normality. This comparison over the course of the frequency function is not satisfactory for small samples unless the grouping is so broad as to deprive the test of most of its efficacy. An alternative is to compute some statistic of the sample and to examine how far it departs from the mean value to be expected on the hypothesis of parent normality.

Consider, for instance, the statistic

$$t = \frac{k_3}{k_2^{3/2}}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (21.14)$$

This is independent of the mean (because the k-statistics are so) and is also independent of the scale parameter because it is " studentised ". In normal samples, therefore, the distribution of t is independent of mean and variance and thus depends only on the sample number n. We have already given formulae for its mean and variance (Exercise 11.16, vol. I, p. 289). In fact,

$$\left.\begin{aligned} \mu_1'(t) &= \mu_3(t) = 0 \\ \mu_2(t) &= \frac{6n(n-1)}{(n-2)(n+1)(n+3)} \end{aligned}\right\} \qquad . \quad . \quad . \quad . \quad (21.15)$$

Since the distribution of t is symmetrical we may, for moderate n, consider it as normally distributed with zero mean and variance given by (21.15), and this will provide a test—of a somewhat approximate kind—of normality in the parent from which the sample is derived.

Example 21.3

In the data of Examples 21.1 and 21.2 we have, for the sample moments about origin 16, in units of 0·1

$$\begin{aligned} m_1' &= -0{\cdot}8 \\ m_2 &= 2{\cdot}16 \\ m_3 &= 0{\cdot}496 \end{aligned}$$

whence
$$k_2 = \frac{n}{n-1} m_2 = 2\cdot31429$$

$$k_3 = \frac{n^2}{(n-1)(n-2)} m_3 = 0\cdot61319$$

and
$$t = \frac{k_3}{k_2^{\frac{3}{2}}} = 0\cdot174.$$

The variance of t, from (21.15), is 0·3188 and its standard error accordingly about 0·57. The observed deviation from zero is considerably less than this, and we see no reason to doubt the hypothesis of normality so far as this test is concerned.

21.19. Another test of normality has been proposed by Geary (1935*a*), namely the use of the ratio

$$w = \frac{\text{mean deviation}}{\text{standard deviation}}. \qquad . \quad . \quad . \quad . \quad . \quad (21.16)$$

If the parent mean is zero, the parent value of w is $\sqrt{\frac{2}{\pi}} = 0\cdot79788$. The test has also been adapted to the case when the parent mean is not zero, and tables provided for the application of the test (Geary and Pearson, 1938).

Geary's ratio is directed towards detecting deviations from mesokurtosis in the parent. The criterion based on k_4/k_2^2, which is a natural extension of that for skewness based on $k_3/k_2^{\frac{3}{2}}$, is not very suitable for the purpose, since it has a skew distribution for quite high values of n. The distribution of Geary's ratio tends to normality fairly rapidly (cf. Exercise 21.2).

Tests of Goodness of Fit

21.20. In Chapter 12 we considered in some detail the use of χ^2 in testing correspondence between observation and hypothesis. If the hypothesis specifies the theoretical values completely no question of estimation arises, and each cell contributing to χ^2 could, if so desired, be tested separately. From this point of view χ^2 compounds into a single test a number of tests of the kind already considered.

If the hypothesis does not specify the theoretical values completely, but leaves them to be estimated in part from the data, some modification in the χ^2-test is necessary. We can now establish a result which in **12.13** was announced without proof: if the estimators employed are maximum likelihood estimators, then for large samples the χ^2-test of significance retains its validity, provided that the number of degrees of freedom is reduced by unity for every parameter estimated.

Suppose the hypothesis leaves unspecified a parameter θ, and let t be its maximum likelihood estimator. Then if the theoretical frequencies based on the true value of θ are λ and those based on t are λ', we may write

$$\chi^2 = \Sigma \frac{(l-\lambda)^2}{\lambda} \qquad . \quad . \quad . \quad . \quad . \quad (21.17)$$

$$\chi'^2 = \Sigma \frac{(l-\lambda')^2}{\lambda'}. \qquad . \quad . \quad . \quad . \quad . \quad (21.18)$$

χ^2 is distributed as the sum of squares of ν normal variates with unit variance. The problem is to find the distribution of χ'^2. We have

$$\chi^2 - \chi'^2 = \Sigma l^2 \left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right),$$

and for large samples the difference between λ and λ' will be of order $n^{-\frac{1}{2}}$. We then have, expanding the difference in terms of $\delta\theta$, to order n^{-1},

$$\frac{1}{\lambda} - \frac{1}{\lambda'} = -\frac{1}{\lambda'^2}\frac{\partial\lambda'}{\partial\theta}\delta\theta + \left\{\frac{2}{\lambda'^3}\left(\frac{\partial\lambda'}{\partial\theta}\right)^2 - \frac{1}{\lambda'^2}\frac{\partial^2\lambda'}{\partial\theta^2}\right\}\frac{(\delta\theta)^2}{2} + \ldots \quad . \quad (21.19)$$

Now for large samples the maximisation of the likelihood is equivalent to minimising χ^2, and hence

$$\Sigma\left(\frac{l^2}{\lambda'^2}\frac{\partial\lambda'}{\partial\theta}\right) = 0,$$

and

$$\chi^2 - \chi'^2 = \frac{(\delta\theta)^2}{2}\Sigma\left\{\frac{2}{\lambda'}\left(\frac{\partial\lambda'}{\partial\theta}\right)^2 - \frac{\partial^2\lambda'}{\partial\theta^2}\right\}$$

$$= (\delta\theta)^2\,\Sigma\left\{\frac{1}{\lambda'}\left(\frac{\partial\lambda'}{\partial\theta}\right)^2\right\}. \quad . \quad . \quad . \quad . \quad . \quad (21.20)$$

But the sum on the right is the reciprocal of the variance of the maximum likelihood estimator, and writing δt for $\delta\theta$, as is legitimate for large samples, we have

$$\chi^2 - \chi'^2 = \frac{(\delta t)^2}{\text{var}\, t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (21.21)$$

The quantity on the right is itself the square of a variate which (in the limit) is normal and has unit variance. Furthermore, its distribution is independent of that of χ'^2. For consider the spherically symmetric density-distribution of the ν normal variables whose sum of squares composes χ^2. Let O be the origin and P any point; then $\chi^2 = OP^2$. Now for large samples the variation takes place in the neighbourhood of O. A surface of constant t through P is approximately plane in the effective range of variation. If OQ is the normal to this surface,

$$OP^2 = OQ^2 + PQ^2,$$

corresponding to

$$\chi^2 = \frac{(\delta t)^2}{\text{var}\, t} + \chi'^2,$$

for t is chosen so as to minimise $\chi'^2 = PQ^2$. Thus if we take t as a new co-ordinate, together with $(\nu - 1)$ others in the surface of constant t, the axis of t is orthogonal to the space of constant t, and t will be independent of χ'^2.

It follows further that χ'^2 is distributed as the sum of $(\nu - 1)$ squares of normal variates. Thus the usual Type III distribution of χ^2 holds for $\nu - 1$ degrees of freedom; and so for every constant fitted, with a reduction of unity in the number of degrees for each constant. We have already exemplified the use of the result in Example 12.4 (Vol. I, p. 301).

The ω^2-distribution

21.21. For small samples the χ^2-test is difficult to apply, since it depends for its validity on the fact that the binomial distribution in individual cells may be represented by the normal distribution, and hence requires that cell-frequencies shall not be small.

A test of a different kind has been proposed by Cramér (1928) and independently by von Mises (1931).

Put

$$\omega^2 = \int_{-\infty}^{\infty} \{ \bar{F}(x) - F(x) \}^2 \, dx, \quad . \quad . \quad . \quad . \quad (21.22)$$

where $\bar{F}(x)$ is the observed *distribution* function and $F(x)$ the hypothetical distribution function. The quantity ω^2 varies from sample to sample, its mean value being

$$E(\omega^2) = \frac{1}{n}\int_{-\infty}^{\infty} F(x)\{1 - F(x)\}\, dx = \frac{1}{2n}\Delta_1, \quad . \quad . \quad . \quad (21.23)$$

where Δ_1 is Gini's coefficient of mean difference (cf. **2.24**). For

$$E(\omega^2) = \int_{-\infty}^{\infty} E\{\bar{F} - F\}^2 \, dx.$$

For any given x the expectation of $(\bar{F} - F)^2$ is merely the variance of the proportion F and hence is equal to $\frac{F(1-F)}{n}$. The result (21.23) follows at once.

The ω^2-test consists of comparing the observed with the mean value; but it is not possible to express the comparison in terms of probability as the sampling distribution of ω^2 is not known.

21.22. The numerical evaluation of the integral (21.22) is tedious in the case of a continuous distribution, and Wold (1938*a*) has suggested a modification. If the variate range is divided into intervals at $-\infty, x_1, x_2 \ldots x_j \ldots \infty$, we define

$$w^2 = \sum_j \{ \bar{F}(x_j) - F(x_j) \}^2 . \quad . \quad . \quad . \quad . \quad (21.24)$$

If the intervals are all of width h,

$$E(w^2) = \frac{1}{nh}\int_{-\infty}^{\infty} F(x)\{1 - F(x)\}\, dx + \frac{1}{n}R, \quad . \quad . \quad . \quad (21.25)$$

where $R/_n$ is a remainder term. If this may be neglected, the w^2-test is equivalent to the ω^2-test but easier to apply. If the data are ungrouped, the x_j's may be taken at equidistant intervals.

In the particular case when F is normal, we have

$$n\,E(\omega^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2} \int_{x}^{\infty} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}v^2}\, du\, dv\, dx. \quad . \quad . \quad (21.26)$$

Putting $u = \alpha + x$ and $v = \beta + x$, we find, after integration with respect to x,

$$\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{0}\int_{0}^{\infty} \exp\{-\tfrac{1}{4}(\alpha - \beta)^2\}\, d\alpha\, d\beta.$$

A further substitution of $\gamma = \alpha - \beta$ and $\delta = \alpha + \beta$ gives

$$\frac{1}{4\sqrt{\pi}}\int_{-\gamma}^{\gamma}\int_{0}^{\infty} e^{-\frac{1}{4}\gamma^2}\, d\gamma\, d\delta$$

or

$$\frac{1}{2\sqrt{\pi}}\int_{0}^{\infty} \gamma\, e^{-\frac{1}{4}\gamma^2}\, d\gamma$$

$$= \frac{1}{\sqrt{\pi}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21.27)$$

21.23. An interesting modification of the ω^2-test has been given by Smirnoff (1936) who defines

$$\omega_n^2 = \int_{-\infty}^{\infty} (\bar{F} - F)^2 \, dF. \qquad . \quad . \quad . \quad . \quad . \quad (21.28)$$

The difference lies in the differential element which has the effect of rendering the distribution of ω_n^2 independent of F. It is shown that as n tends to infinity the distribution function of ω_n^2 tends to the form

$$1 - \frac{1}{\pi} \sum_{k=1}^{\infty} \int_{(2k-1)n}^{2kn} \frac{e^{-\frac{1}{2}z^2\omega^2_n} \, dz}{\sqrt{(-z \sin z)}}, \qquad . \quad . \quad . \quad . \quad . \quad (21.29)$$

but this does not look a very promising formula for application in particular cases.

Cramér (1928) has extended formula (21.27) to the goodness of fit of Gram-Charlier series and gives some examples of fitting to observed distributions.

Difference of Two Means

21.24. A common case occurring in practice is that of two independent samples of n_1 and n_2 members from two populations which may or may not be different. We wish to decide whether the evidence indicates a significant difference between the parent means. This situation forms a kind of border-line case between the testing of a prior value of a parameter and the homogeneity tests which we shall consider below. It is a test of homogeneity in the sense that we are to discuss the question whether two populations are equal in certain respects; but we do not necessarily assume that they are identical, and in any case we can regard the problem as equivalent to the testing of a single parameter (the difference of the means) to see whether it is different from zero.

21.25. For large samples we discussed the question in Example 9.10 (Vol. I, p. 226) and gave two tests. If the hypothesis is that the parent populations are identical (a true hypothesis of homogeneity) we may pool the samples to form a single sample and test whether either mean differs from the mean of the total. If, however, we wish to test the less general hypothesis that the parents have the same mean but not necessarily the same variance, we may test the difference of means by the ordinary equation expressing the variance of a difference in terms of the separate variances. This is not a homogeneity test in the strictest sense of the word, but tests of such a character may conveniently be discussed in conjunction with the other type, both for small and for large samples.

21.26. We now consider the corresponding problem when the samples are small and the parent populations are assumed to be normal. In the first place we take the case when the two populations have the same variance σ^2.

The sample means $\bar{x}_1$ and $\bar{x}_2$ are distributed normally with variances $\frac{\sigma^2}{n_1}$ and $\frac{\sigma^2}{n_2}$ and means μ_1 and μ_2. Consequently $\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma}$ is distributed normally with variance $\frac{1}{n_1} + \frac{1}{n_2}$, and hence

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \qquad . \quad . \quad . \quad . \quad . \quad (21.30)$$

is distributed normally with unit variance about zero mean. Further, if S_1^2 and S_2^2 are the sample sums of squares about the mean, the quantity

$$\frac{S^2}{\sigma^2} = \frac{1}{\sigma^2}(S_1^2 + S_2^2) \qquad . \quad . \quad . \quad . \quad . \quad (21.31)$$

is distributed as χ^2 with $n_1 + n_2 - 2$ degrees of freedom, independently of the expression (21.30). It follows that

$$u = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{S} \sqrt{\left\{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}\right\}} \qquad . \quad . \quad (21.32)$$

is distributed like "Student's" t with $\nu = n_1 + n_2 - 2$ degrees of freedom. This expression does not contain the unknown σ and hence may be used to test the difference $\mu_1 - \mu_2$. This result is due to Fisher (1926*a*).

Example 21.4

In a class of 20 children, 10 chosen at random were given a ration of orange-juice each day for a certain period and the other 10 a ration of milk. Their gains in weight during the period were, in pounds :—

First group : 4, $2\frac{1}{2}$, $3\frac{1}{2}$, 4, $1\frac{1}{2}$, 1, $3\frac{1}{2}$, 3, $2\frac{1}{2}$, $3\frac{1}{2}$
Second group : $1\frac{1}{2}$, $3\frac{1}{2}$, $2\frac{1}{2}$, 3, $2\frac{1}{2}$, 2, 2, $2\frac{1}{2}$, $1\frac{1}{2}$, 3

The mean increase in the first group is 2·9 pounds, and in the second 2·4 pounds. Putting aside other explanations, one possible factor accounting for this difference is the difference in treatments. But we wish to know in the first place whether this is significant. We assume, then, that treatment exerted no differential effect and that the samples came from normal populations with the same mean and variance. We find

$$\bar{x}_1 = 2{\cdot}9 \qquad\qquad \bar{x}_2 = 2{\cdot}4$$
$$\Sigma(x_1 - \bar{x}_1)^2 = 9{\cdot}4 \qquad \Sigma(x_2 - \bar{x}_2)^2 = 3{\cdot}9.$$

Hence, from (21.32), with $\mu_1 - \mu_2 = 0$,

$$\nu = 10 + 10 - 2 = 18$$
$$u = \frac{0{\cdot}5}{\sqrt{13{\cdot}3}}\sqrt{18}\sqrt{\frac{100}{20}} = 1{\cdot}30.$$

From Appendix Table 3 (vol. I, p. 441) we see that such a value would be exceeded in absolute value with probability 0·21. The difference of a half-pound between the sample means is not significant.

We note incidentally that the sample variances, 0·940 and 0·390, differ considerably, and shall see below how the significance of the difference may be tested. At the present stage our conclusion as to the non-significance of the difference of means is to be regarded with reserve, for the data themselves suggest that we have over-simplified the problem in assuming equal variance in the two populations.

21.27. Apart from the question of unequal variances, the data of the previous example will serve to illustrate a further point of interest. Our hypothesis is that the children within each group may be regarded as a sample from a population with the same mean. Had we been dealing with a sample of, say, seedlings grown from the seed of a single plant, this hypothesis would not have been unreasonable ; but children differ very much among themselves in nutritional standard, and so forth. Our hypothesis is again liable to over-simplify the problem.

When the statistician can direct the sampling himself, this kind of problem can be tackled with success by pairing. Suppose we select children in pairs of the same sex, each pair resembling each other as closely as possible in all the factors which might influence the experiment such as age, weight and nutritional standard. We allot at random one member to the first group and one to the second, and so for each pair. The differences in weights gained between members of a pair may then be regarded as samples from a population with zero mean, even if the pairs differ among themselves, and the set of differences tested in the usual way.

Example 21.5

Suppose that, in the previous example, the data had related to 10 pairs of children, thus :—

No. of Pair.	First Group wt. in lbs.	Second Group wt. in lbs.	Difference, First – Second.
1	4	$1\frac{1}{2}$	$2\frac{1}{2}$
2	$2\frac{1}{2}$	$3\frac{1}{2}$	-1
3	$3\frac{1}{2}$	$2\frac{1}{2}$	1
4	4	3	1
5	$1\frac{1}{2}$	$2\frac{1}{2}$	-1
6	1	2	-1
7	$3\frac{1}{2}$	2	$1\frac{1}{2}$
8	3	$2\frac{1}{2}$	$\frac{1}{2}$
9	$2\frac{1}{2}$	$1\frac{1}{2}$	1
10	$3\frac{1}{2}$	3	$\frac{1}{2}$
TOTALS	29	24	5

For the values in the last column we find

$$\bar{x} = 0\cdot5 \qquad s^2 = 1\cdot25 \qquad \nu = 9$$

$$t = \frac{0\cdot5}{\sqrt{1\cdot25}}\sqrt{9} = 1\cdot34.$$

The probability of obtaining such a value or greater (absolutely) is about 0·22, and the observed differences are therefore not significant. This is the same conclusion that we reached in Example 21.3, but it would not have been surprising had the conclusions differed, for they relate to different questions.

Difference of Means when Variances are Unequal

21.28. When population variances are not assumed equal the t-test of difference of means no longer applies. We can, if we choose, apply a test based on fiducial intervals, namely, the Behrens test, considered in the previous chapter. We put

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_2'^2 + s_2'^2)}}. \qquad . \quad . \quad . \quad . \quad (21.33)$$

The fiducial limits of d for various significance levels have been tabulated by Sukhatme

(1938*b*) and Fisher (1941*a*) for n_1 and n_2 greater than 5. If the observed d falls inside the range, we may accept the hypothesis that the population means are equal.

21.29. As we have seen, an inference of this kind does not imply that we shall be correct in a certain proportion of the cases, and if we wish to find a test satisfying such a criterion a different approach is necessary. The following investigation is due to Welch (1938*b*).

Consider the distribution of u of equation (21.32) when the means are the same but the variances are different, i.e.

$$u = \frac{\bar{x}_1 - \bar{x}_2}{\left\{\frac{S_1^2 + S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{\frac{1}{2}}}. \qquad (21.34)$$

Put

$$\bar{x}_1 - \bar{x}_2 = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}} \chi' . \qquad (21.35)$$

$$w = \frac{\sigma_1^2 \chi_1^2 + \sigma_2^2 \chi_2^2}{(n_1 + n_2 - 2)\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right), \qquad (21.36)$$

where $\sigma_1^2 \chi_1^2 = S_1^2$ and hence χ_1^2 is distributed as χ^2 with $\nu_1 = n_1 - 1$ degrees of freedom, and similarly for χ_2^2. χ' may be regarded as a single normal variate with zero mean and unit variance. We have then

$$u = \frac{\chi'}{\sqrt{w}}. \qquad (21.37)$$

Now put

$$w = a\chi_1^2 + b\chi_2^2, \qquad (21.38)$$

where, from (21.36),

$$\left.\begin{aligned} a &= \frac{\sigma_1^2}{n_1 + n_2 - 2}\,\frac{\frac{1}{n_1} + \frac{1}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\ b &= \frac{\sigma_2^2}{n_1 + n_2 - 2}\,\frac{\frac{1}{n_1} + \frac{1}{n_2}}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \end{aligned}\right\} \qquad (21.39)$$

w itself is not distributed in the Type III form unless $\sigma_1 = \sigma_2$, but we will find a distribution of that form which approximates to it by equating lower moments. The first two moments of w, being the sum of the separate parts, are

$$\left.\begin{aligned} \mu_1'(w) &= a\nu_1 + b\nu_2 \\ \mu_2(w) &= 2(a^2\nu_1 + b^2\nu_2) \end{aligned}\right\}. \qquad (21.40)$$

The moments of

$$dF = \frac{1}{(2g)^{\frac{1}{2}\nu}\,\Gamma(\frac{1}{2}\nu)}\, w^{\frac{1}{2}\nu - 1}\, e^{-w/2g}\, dw$$

are

$$\left.\begin{aligned} \mu_1' &= g\nu \\ \mu_2 &= 2g^2\nu \end{aligned}\right\}. \qquad (21.41)$$

Identifying (21.40) and (21.41) we find—

$$\left.\begin{aligned} g &= \frac{a^2 \nu_1 + b^2 \nu_2}{a\nu_1 + b\nu_2} \\ \nu &= \frac{(a\nu_1 + b\nu_2)^2}{a^2 \nu_1 + b^2 \nu_2} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (21.42)$$

With these values of g and ν the distribution of w/g is approximately of the Type III form with ν degrees of freedom and will be independent of χ'. Hence,

$$\frac{\chi' \sqrt{\nu}}{\sqrt{\dfrac{w}{g}}} = \chi' \sqrt{\frac{g\nu}{w}}$$

$$= u\sqrt{(g\nu)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21.43)$$

is distributed approximately as " Student's " t with ν degrees of freedom. In particular, if $\sigma_1 = \sigma_2$, $a = b$ and we reduce to the test of **21.26.**

21.30. In general, when $\sigma_1 \neq \sigma_2$ the quantities g and ν depend on the ratio $\theta = \sigma_1^2/\sigma_2^2$. We have

$$\nu = \frac{(\nu_1 \theta + \nu_2)^2}{\nu_1 \theta^2 + \nu_2} \quad . \quad . \quad . \quad . \quad . \quad (21.44)$$

and may put $u = ct$ where $c = 1/\sqrt{\nu g}$, and hence

$$c = \left\{ \frac{(\nu_1 + \nu_2)\left(\dfrac{\theta}{n_1} + \dfrac{1}{n_2}\right)}{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)(\nu_1 \theta + \nu_2)} \right\}^{\frac{1}{2}} \quad . \quad . \quad . \quad . \quad (21.45)$$

Without a definite knowledge of θ we cannot apply the t-test, but the advantage of putting the expressions in this form is that by considering particular values of θ we are able to judge how far the test based on " Student's " distribution is likely to be affected.

Example 21.6 (from Welch, 1938*b*)

Consider the case $n_1 = n_2 = 10$. From (21.45) we have $c = 1$ and from (21.44)

$$\nu = \frac{9\,(\theta + 1)^2}{\theta^2 + 1}.$$

Suppose now we were to use the test of **21.26**, based on the assumption that $\theta = 1$. We should find, to a probability level of 0·05, that $|u|$ must exceed 2·101 to be significant. If we judge u significant for such values how far are we in error when θ is not unity ? That is to say, what are the true probabilities that

$$P\{|u| > 2{\cdot}101\}$$

for varying values of θ, as compared with our value of 0·05 ?

For a specified θ the probabilities can easily be obtained from the approximate distribution $u\sqrt{(g\nu)}$ of equation (21.43). They are shown graphically in Fig. 21.1. The full line (a) shows P for various values of θ and $n_1 = n_2 = 10$. The full line (b) shows similarly the values for $n_1 = 5$, $n_2 = 15$. (The dotted line (c) we refer to below.)

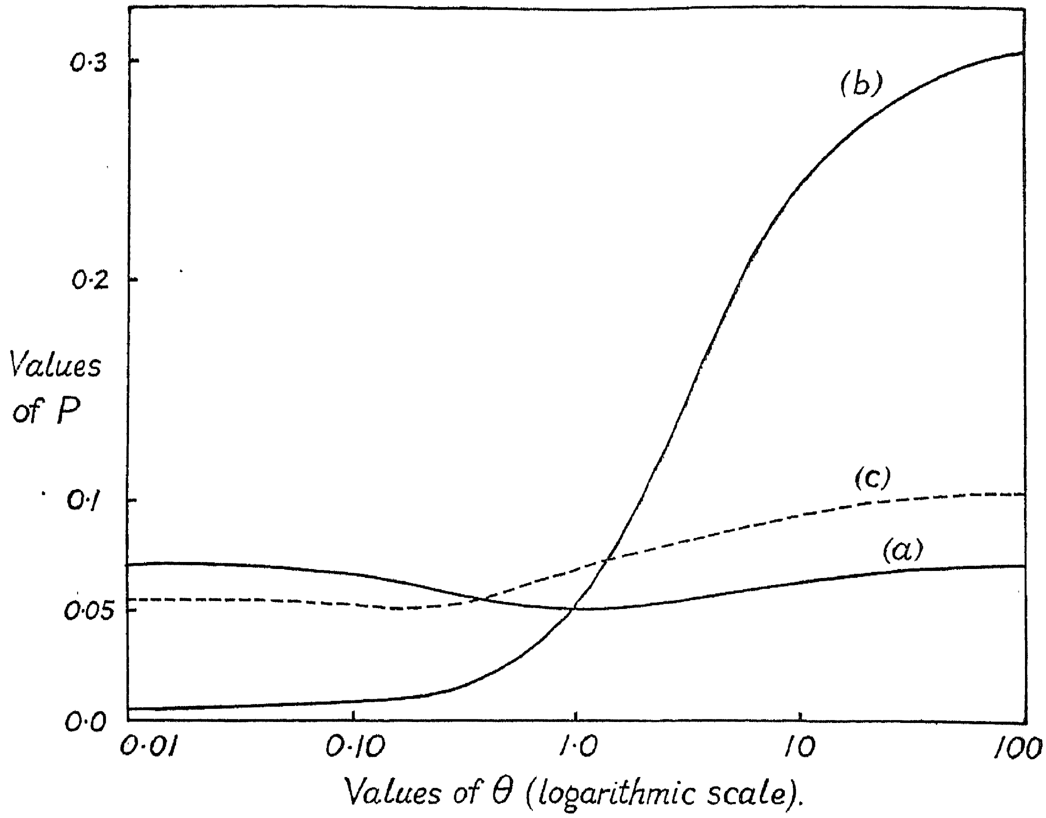


FIG. 21.1.

In case (a) the line does not deviate very much from the horizontal at $P = 0{\cdot}05$, and we may conclude that the test based on the assumption of equal variance is not very much in error. In any case, if the curve falls below the line $P = 0{\cdot}05$ we are on the safe side, for our true probability is then less than 0·05, and in rejecting the hypothesis at that level we are adopting more stringent standards than is apparent.

In case (b), when the sample numbers are unequal we have a different state of affairs. For $\theta < 1$ the test is very conservative, but for $\theta > 1$ it may err very seriously in the wrong direction.

21.31. Welch concludes that for samples of equal size there is not a serious likelihood of error in testing the difference of means as if the parent variances were equal. For samples of unequal size the error may invalidate the t-test and an alternative criterion is proposed. Write

$$\frac{\bar{x}_1 - \bar{x}_2}{\left\{\dfrac{S_1^2}{n_1(n_1-1)} + \dfrac{S_2^2}{n_2(n_2-1)}\right\}^{\frac{1}{2}}}. \qquad (21.46)$$

Here, it will be observed, the denominator is an estimate of $\left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}$, the standard deviation of the difference $\bar{x}_1 - \bar{x}_2$. Precisely as for u we approximate to the distribution of this denominator by a Type III form. Corresponding to (21.39) we find

$$\left.\begin{aligned} a &= \frac{\sigma_1^2}{n_1(n_1-1)} \Big/ \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \\ b &= \frac{\sigma_2^2}{n_2(n_2-1)} \Big/ \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \end{aligned}\right\} \qquad (21.47)$$

Corresponding to (21.45) we find $c = 1$, and to (21.44)

$$\nu = \left(\frac{\theta}{n_1} + \frac{1}{n_2}\right)^2 \Big/ \left(\frac{\theta_1^2}{n_1^2(n_1-1)} + \frac{1}{n_2^2(n_2-1)}\right). \qquad (21.48)$$

v is then distributed approximately in "Student's" form with ν degrees of freedom. The dotted line (c) in Fig. 21.1 shows the relationship between θ and $P\{|v| > 2{\cdot}101\}$ for $n_1 = 5$, $n_2 = 15$. Clearly the error is now much smaller than when we used u for the same sample numbers.

Difference of Two Variances in Normal Samples

21.32. If we have samples of n_1 and n_2 members from normal populations with variances σ_1^2 and σ_2^2, the ratio of sample variances $p^2 = \frac{s_1^2}{s_2^2}$ is distributed in the form (cf. Example 10.18, vol. I, p. 249)—

$$dF \propto \frac{p^{n_1-2}\,dp}{\left(\frac{n_1 p^2}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}\right)^{\frac{1}{2}(n_1+n_2-2)}}. \qquad (21.49)$$

The related quantity

$$z = \tfrac{1}{2}\log \frac{n_1(n_2-1)}{n_2(n_1-1)} p^2 \qquad (21.50)$$

is distributed in Fisher's form

$$dF \propto \frac{e^{\nu_1 z}\,dz}{\left(\frac{\nu_1 e^{2z}}{\sigma_1^2} + \frac{\nu_2}{\sigma_2^2}\right)^{\frac{1}{2}(\nu_1+\nu_2)}} \qquad (21.51)$$

where $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$. The ν's may, by a convenient extension of our previous terminology, be called the degrees of freedom associated with z. In practice, z is generally used in preference to p, but tables of both are available.

These distributions provide a test of significance of the equality of the ratio σ_1^2/σ_2^2. On the hypothesis of equality they are independent of the ratio and the probability of an observed p or z can be obtained. As usual, if this is small we reject the hypothesis. We leave it to the reader to show that this type of inference can be based on the theory of confidence intervals or the theory of fiducial intervals in the usual way.

Example 21.7

In Example 21.4 we had two samples of children and found that the difference in means was not significant. This was on the hypothesis that the variances were identical, and since the two samples are equal in number the inference remains valid even if the variances are different, as illustrated in **21.31.** We will now test directly whether the sample variances themselves indicate any significant difference in parent variances.

We have

$$\Sigma(x_1 - \bar{x}_1)^2 = 9{\cdot}40 \qquad \nu_1 = 9$$
$$\Sigma(x_2 - \bar{x}_2)^2 = 3{\cdot}90 \qquad \nu_2 = 9.$$

Hence

$$z = \tfrac{1}{2}\log_e \frac{9{\cdot}40}{9} \Big/ \frac{3{\cdot}90}{9} = 0{\cdot}4398.$$

From Appendix Tables 4 and 5 of Vol. I (pp. 442–3) we see that for $\nu_2 = 9$ the 5-per-cent points of z are

$$\nu_1 = 8, \quad 0{\cdot}5862$$
$$\nu_1 = 12, \quad 0{\cdot}5613$$

and the 1-per-cent. points are

$$\nu_1 = 8, \quad 0{\cdot}8494$$
$$\nu_1 = 12, \quad 0{\cdot}8157.$$

Thus, notwithstanding that one variance is about $2\frac{1}{2}$ times the other, the probability that the observed z will be exceeded on random sampling from populations with the same variance is greater than 0·05, and the difference of sample variances is not significant.

There is a point here which is frequently overlooked. In carrying out the z-test we always take the ratio of the larger variance to the smaller, so that our probability levels relate, not to the chance that a given pair of variances have a larger ratio than the observed one, but to the chance that the bigger of the two exceeds the smaller in a certain ratio. A probability of 0·05 thus relates to the chance that *either* s_1^2/s_2^2 exceeds a given amount k, *or* s_1^2/s_2^2 falls short of a given amount $1/k$. If we are interested only in the former contingency our probabilities should be halved.

Properties of Fisher's Distribution

21.33. The z-distribution plays a very important part in statistical inference based on small samples, and we digress at this point to give an account of its main features.

The distribution function of z may be obtained from the incomplete B-function, for z may be easily transformed into a Type I variate. There are, however, special tables for lower values of ν_1 and ν_2 and satisfactory approximations of various kinds for higher values.

The characteristic function of z is proportional to

$$\int_{-\infty}^{\infty} \frac{e^{(\theta+\nu_1)z}\, dz}{(\nu_1 e^{2z} + \nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}$$

where $\theta = it$, and is thus

$$\phi(t) = \left(\frac{\nu_2}{\nu_1}\right)^{\frac{1}{2}\theta} \frac{\Gamma\left(\frac{\nu_2-\theta}{2}\right)\Gamma\left(\frac{\nu_1+\theta}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \quad . \quad . \quad . \quad . \quad (21.52)$$

Thus, taking logarithms and using the expansion

$$\log \Gamma(1+x) = \tfrac{1}{2}\log 2\pi + (x+\tfrac{1}{2})\log x - x + \frac{1}{12x} - \ldots,$$

we find

$$\log \phi(t) = -\frac{\theta}{2}\left(\frac{1}{\nu_1} - \frac{1}{\nu_2}\right) + \frac{\theta^2}{4}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right) - \ldots \quad . \quad . \quad (21.53)$$

Thus, for large ν_1 and ν_2, z is distributed normally with mean

$$-\tfrac{1}{2}\left(\frac{1}{\nu_1} - \frac{1}{\nu_2}\right) \text{ and variance } \tfrac{1}{2}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right).$$

21.34. Various approximations have been given for the case when ν_1 and ν_2 are not large enough to justify the assumption of normality.

(*a*) (Cornish and Fisher, 1937). The method is that of **6.32** and depends on the expansion of the distribution in a Gram-Charlier series. From the successive derivatives of $\log \Gamma(1+x)$ we can find those of $\phi(t)$, and hence ascertain the cumulants of z. Writing $r_1 = \frac{1}{\nu_1}$ and $r_2 = \frac{1}{\nu_2}$, we find

$$\left.\begin{aligned}
\kappa_1 &= -\tfrac{1}{2}(r_1 - r_2) - \tfrac{1}{6}(r_1^2 - r_2^2)\\
\kappa_2 &= \tfrac{1}{2}(r_1 + r_2) + \tfrac{1}{2}(r_1^2 + r_2^2) + \tfrac{1}{3}(r_1^3 + r_2^3)\\
\kappa_3 &= -\tfrac{1}{2}(r_1^2 - r_2^2) - (r_1^3 - r_2^3)\\
\kappa_4 &= r_1^3 + r_2^3 + 3(r_1^4 + r_2^4)\\
\kappa_5 &= -3(r_1^4 - r_2^4)\\
\kappa_6 &= 12(r_1^5 + r_2^5)
\end{aligned}\right\} \quad . \quad . \quad . \quad (21.54)$$

Hence, putting $\sigma = r_1 + r_2$ and $\delta = r_1 - r_2$, we find for the l's of **6.32** ($m = 0$, variance $= \frac{1}{2}\sigma$)—

$$l_1 = -\sqrt{\frac{2}{\sigma}}\left(\tfrac{1}{2}\delta + \tfrac{1}{6}\delta\sigma\right)$$

$$l_2 = \tfrac{1}{2}\left(\sigma + \frac{\delta^2}{\sigma}\right) + \tfrac{1}{6}(\sigma^2 + 3\delta^2),$$

and so on. After some reduction we find, for the value of z corresponding to a probability α (which in turn corresponds to a normal deviate ξ),—

$$z = \xi\sqrt{\frac{\sigma}{2}} - \tfrac{1}{6}\delta(\xi^2 + 2) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma}{24}(\xi^3 + 3\xi) + \frac{1}{72}\frac{\delta^2}{\sigma}(\xi^3 + 11\xi)\right\}$$
$$- \frac{\delta\sigma}{120}(\xi^4 + 9\xi^2 + 8) + \frac{\delta^3}{3240\sigma}(3\xi^4 + 7\xi^2 + 16) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma^2}{1920}(\xi^5 + 20\xi^3 + 15\xi)\right.$$
$$\left. + \frac{\delta^4}{2880}(\xi^5 + 44\xi^3 + 183\xi) + \frac{\delta^4}{155520\sigma^2}(9\xi^5 - 284\xi^3 - 1513\xi)\right\} \quad . \quad (21.55)$$

(*b*) (Fisher, extended by Cochran, 1940*a*). Writing n indifferently for ν_1 and ν_2, we have, from (21.55), to order n^{-2}—

$$z = \xi\sqrt{\frac{\sigma}{2}} - \tfrac{1}{6}\delta(\xi^2 + 2) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma}{24}(\xi^3 + 3\xi) + \frac{1}{72}\frac{\delta^2}{\sigma}(\xi^3 + 11\xi)\right\}.$$

Put $h = 2/\sigma$. Then

$$z = \frac{\xi}{\sqrt{h}} - \tfrac{1}{6}\delta(\xi^2 + 2) + \frac{1}{\sqrt{h}}\left\{\frac{\xi^3 + 3\xi}{12h} + \frac{\xi^3 + 11\xi}{144}h\delta^2\right\}. \quad . \quad (21.56)$$

Now
$$\frac{\xi}{\sqrt{(h - \lambda)}} = \frac{\xi}{\sqrt{h}} + \frac{\lambda\xi}{2h\sqrt{h}} + O(n^{-2}).$$

Hence, if we put

$$z = \frac{\xi}{\sqrt{(h - \lambda)}} - \tfrac{1}{6}\delta(\xi^2 + 2), \quad . \quad . \quad . \quad . \quad (21.57)$$

the difference of this quantity from (21.56) is

$$\frac{(\xi^3 + 11\xi)\,\delta^2\sqrt{h}}{144},$$

provided that we take $\lambda = \dfrac{\xi^2 + 3}{6}$.

The difference is small in virtue of the large denominator and the factor $\delta^2 = \left(\dfrac{1}{\nu_1} - \dfrac{1}{\nu_2}\right)^2$ which is small if ν_1 and ν_2 are not too different. Thus we may take z as approximately given by (21.57). The values of λ for various values of the significance level are

Level	40%	30%	20%	10%	5%	1%	0·1%
λ	0·51	0·55	0·62	0·77	0·95	1·40	2·09

For the commoner levels of significance the form taken by (21.57) is

$$20 \text{ per cent. level}: \quad \frac{0{\cdot}8416}{\sqrt{(h-\lambda)}} - 0{\cdot}4514\delta \quad . \quad . \quad . \quad (21.58)$$

$$5 \text{ per cent. level}: \quad \frac{1{\cdot}6449}{\sqrt{(h-\lambda)}} - 0{\cdot}7843\delta \quad . \quad . \quad . \quad (21.59)$$

$$1 \text{ per cent. level}: \quad \frac{2{\cdot}3263}{\sqrt{(h-\lambda)}} - 1{\cdot}235\delta. \quad . \quad . \quad . \quad (21.60)$$

$$0{\cdot}1 \text{ per cent. level}: \quad \frac{3{\cdot}0902}{\sqrt{(h-\lambda)}} - 1{\cdot}925\delta. \quad . \quad . \quad . \quad (21.61)$$

The accuracy of the approximation for $\nu_1 = 24$, $\nu_2 = 60$ may be judged from the following comparison :—

Level per cent.	Value of z from (21.57).	Exact Value.
20	0·1337	0·1338
1	0·3748	0·3746
0·1	0·4966	0·4955

(*c*) (Paulson, 1942). The Wilson–Hilferty approximation to χ^2 of **12.7** indicates that $\left(\frac{\chi^2}{\nu}\right)^{\frac{1}{3}}$ is distributed normally about mean $1 - \frac{2}{9\nu}$ with variance $\frac{2}{9\nu}$. The ratio $\frac{s_1^2}{s_2^2}$ itself is the ratio of two independent quantities distributed as χ^2 with ν_1 and ν_2 degrees of freedom. Further, in virtue of Geary's theorem (Vol. I, p. 253) the ratio $\frac{m_1 - m_2 p}{(\sigma_1^2 + \sigma_2^2 p^2)^{\frac{1}{2}}}$ is normally distributed in standard measure.

We may thus regard

$$u = \frac{\left(1 - \frac{2}{9\nu_2}\right)\left(\frac{s_1}{s_2}\right)^{\frac{2}{3}} - \left(1 - \frac{2}{9\nu_1}\right)}{\left\{\frac{2}{9\nu_2}\left(\frac{s_1}{s_2}\right)^{\frac{4}{3}} + \frac{2}{9\nu_1}\right\}^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (21.62)$$

as approximately normally distributed in standard measure. The approximation seems remarkably good. For instance, the following shows the exact and approximate values of p^2 for $\nu_1 = 6$, $\nu_2 = 12$.

Level per cent.	$\left(\frac{s_1}{s_2}\right)^2 = p^2$, from (21.62).	Exact Value.
20	1·72	1·72
5	3·00	3·00
1	4·85	4·82
0·1	8·58	8·38

The Problem of k Samples

21.35. We now proceed to consider the case when we have samples from k different populations and wish to determine whether there is any evidence of significant differences between those populations. In some cases the appropriate test can be carried out by the χ^2-distribution, particularly if the data are grouped. For the groups may then be regarded as determining the rows of a contingency table and the different samples the columns, and a homogeneity test applied to the table in the manner of Chapter 12. Again, we may compare the samples pair by pair by the foregoing methods ; but this, apart from being tedious, does not give us what we want, namely a test of homogeneity of the set of samples taken together.

21.36. Consider in the first instance the sampling of attributes. Suppose we have samples from populations in which the true proportions of successes are ϖ, the observed proportions being $p_1 \ldots p_k$ and the sample numbers $n_1 \ldots n_k$, totalling n.

If p is the mean proportion of successes in all samples taken together, and our hypothesis is that the populations have a common value, p will be an estimate of ϖ and we have for the variance of p_j—

$$\operatorname{var} p_j = \frac{\varpi\chi}{n_j}$$

$$= \frac{pq}{n_j} \text{ approximately,} \qquad (21.63)$$

where

$$p = \frac{1}{n} \Sigma n_j p_j.$$

It follows that $(p_j - p)\sqrt{\dfrac{n_j}{pq}}$ will be distributed normally about zero mean with unit variance, and hence

$$\chi^2 = \frac{\Sigma \{ n_j (p_j - p)^2 \}}{pq} \qquad (21.64)$$

in the Type III form with $k - 1$ degrees of freedom (not k because we have lost a degree by estimating p). Hence the ratio

$$Q^2 = \frac{\Sigma n_j (p_j - p)^2}{pq (k - 1)} \qquad (21.65)$$

has expectation unity. The quantity Q is called the Lexis ratio, after the author who first discussed it in detail (Lexis, 1903).*

* Lexis first developed the use of Q in a paper " Über die Theorie der Stabilität statistischer Reihen," 1879, *Conrad's Jahrbücher*, **32**, 60, reproduced in the reference given above. He dealt, however, only with the case when all the n's were equal and had no knowledge of the sampling distribution of Q. In practical applications he took as each n_j the average for the group. " Der dadurch begangenen Fehler kann man beurteilen wennman n einmal mit der grössten und einmal mit der kleinsten Grundzahl berechnet."

Example 21.8

From 1910 to 1919 the numbers of live male and female births in England and Wales were as follows :—

Year.	Male Births.	Female Births.	Total Births.	Proportion Male/Total.
1910	457,266	439,696	896,962	0·5098
1911	448,933	432,205	881,138	0·5095
1912	445,004	427,733	872,737	0·5099
1913	449,159	432,731	881,890	0·5093
1914	447,184	431,912	879,096	0·5087
1915	415,205	399,409	814,614	0·5097
1916	402,137	383,383	785,520	0·5119
1917	341,361	326,985	668,346	0·5108
1918	339,112	323,549	662,661	0·5117
1919	356,241	336,197	692,438	0·5145
TOTALS	4,101,602	3,933,800	8,035,402	0·5104

The proportion of male births showed an increase during the war years 1916–1919. This is a well-known effect of war, but suppose we had noticed it here for the first time. The natural question is : can the effect be accidental ? There is no doubt about its *reality*, for the data cover the whole population ; but if we suppose that sex at birth is distributed according to the laws of chance, do the differences observed suggest that in the ten years concerned there was a significant change in the population (as regards proportion of male births) ? Let us consider the homogeneity test applied to the 10 proportions.

We have $p = 0{\cdot}5104$, $n = 8{,}035{,}402$, $k - 1 = \nu = 9$ and the sum $\Sigma n_j (p_j - p)^2$ will be found to be 19·895,783. Hence

$$Q = \sqrt{\frac{19{\cdot}895{,}783}{9 \times 0{\cdot}5104 \times 0{\cdot}4896}} = 2{\cdot}974$$

$$\chi^2 = (k - 1)\, Q^2 = 79{\cdot}618.$$

Q is sufficiently far from unity to reject decisively the hypothesis that the data are homogeneous. A χ^2-test will confirm the conclusion. We infer that, whatever the reason, the differences in proportions of male births, slight as they are, cannot be accounted for on the supposition that the distribution of sex is according to chance in samples from a constant population. We may observe that, had we obtained the same proportions for a sample one-tenth the size, χ^2 would have been 7·962 and we should not have inferred non-homogeneity.

21.37. A similar test may be applied with k samples of variables. Let the samples be

$$\begin{array}{llll} x_{11}, & x_{12}, & \ldots & x_{1n_1} \text{ with mean } \bar{x}_1 \\ x_{21}, & x_{22}, & \ldots & x_{2n_2} \quad ,, \quad ,, \quad \bar{x}_2 \\ \cdot & \cdot & \cdot & \cdot \\ x_{k1}, & x_{k2}, & \ldots & x_{kn_k} \quad ,, \quad ,, \quad \bar{x}_k. \end{array}$$

The variance of the jth sample is

$$\frac{1}{n_j} \sum_{l=1}^{n_j} (x_{jl} - \bar{x}_j)^2,$$

and an estimate of the population variance may be obtained by taking the *weighted* mean of sample variances

$$s_v^2 = \frac{1}{n-k} \sum_j \sum_l (x_{jl} - \bar{x}_j)^2. \quad . \quad . \quad . \quad . \quad (21.66)$$

Here we have reduced the divisor to $n - k$ so as to correspond with the number of degrees of freedom.

Furthermore $\bar{x}_j$ will be distributed with variance $\frac{\sigma^2}{n_j}$ and hence (assuming without loss of generality that the parent mean is zero),

$$E \sum_{j=1}^{k} \{n_j (\bar{x}_j - \bar{x})^2\} = \Sigma \{E (n_j \bar{x}_j^2) - E (n\bar{x}^2)\}$$
$$= k\sigma^2 - \sigma^2$$
$$= (k - 1)\sigma^2.$$

Putting then

$$s_u^2 = \frac{1}{k-1} \sum_j n_j (\bar{x}_j - \bar{x})^2, \quad . \quad . \quad . \quad . \quad (21.67)$$

we have another estimate of σ^2. Within sampling limits s_v and s_u should be equal. If they are not, we suspect the homogeneity of the population.

21.38. The above test is a simple form of the analysis of variance, which we shall study extensively in Chapters 23 and 24; it is therefore unnecessary for us to develop it further at the present stage. Essentially the test is one of simultaneous significance of differences between *means* on the assumption that variances are constant. We shall also discuss in Chapter 26 a generalisation of the variance ratio for testing the homogeneity of a set of *variances*.

Example 21.9

The following table (from the Registrar-General's *Statistical Review of England and Wales for 1933*, Part II) shows the numbers of males married in England in that year classified according to age and district. (Certain small numbers of unspecified age and those under 21 have been omitted.)

District.	Age (Years). 21–	25–	30–	35–	45–	55–	Totals.
South-East . .	31,714	43,979	14,995	7,985	3,928	3,717	106,318
North. . . .	31,507	39,849	13,620	7,108	3,362	2,916	98,362
Midland . . .	17,465	21,486	6,729	3,340	1,624	1,509	52,153
East	4,016	5,297	1,820	962	457	386	12,938
South-West . .	4,323	6,065	2,218	1,177	514	580	14,877
Totals	89,025	116,676	39,382	20,572	9,885	9,108	284,648

Note the changes in interval at 25– and 35– years.

The question we shall consider is whether age at marriage differs significantly between different districts. This might, for example, be an important point if we were about to sample the population for some quality related to age at marriage, such as the number of children per family. The data might be regarded as a contingency table and χ^2 used as a test of independence in the usual way. Here we adopt an alternative by considering the mean age at marriage in the five different districts.

Taking the centres of the intervals to be 23, 27·5, 32·5, 40, 50 and 57·5 years (the latter being admittedly an approximation) and making no corrections for grouping, we find :—

District.	Number.	Mean (years).	Sum of Squares of Deviations from Mean.	Variance.
South-East	106,318	29·681,799	7,092,490	66·710
North	98,362	29·312,626	6,092,375	61·938
Midland	52,153	29·007,344	3,105,520	59·546
East	12,938	29·425,761	807,911	62·445
South-West.	14,877	29·873,731	1,025,284	68·917
Whole population . . .	284,648	29·429,049	18,143,921	63·741

The total of the sum of squares about district means, $\Sigma\,(x_{jl} - \bar{x}_j)^2$, is the sum of the figures in the fourth column, namely 18,123,580. The sum of squares $\Sigma\, n_j\,(\bar{x}_j - \bar{x})^2$ is found to be 20,341. We have the useful check that these two together are equal to the sum of squares of deviations from the population mean, 18,143,921 (a property which we shall often require in the analysis of variance).

Thus

$$s_v^2 = \frac{18{,}123{,}580}{284{,}648} = 63{\cdot}67$$

$$s_u^2 = \frac{20{,}341}{4} = 5085{\cdot}25.$$

No test of significance is required to see that the difference in mean age at marriage between districts is not a chance effect.

Tests of Random Order

21.39. The tests described above are concerned with the values of a number of sample members but not with the order in which these values occur. Sometimes there may not be an order, as, for instance, if a number of plants are grown simultaneously or a number of names drawn from a hat in a single handful. More frequently there is a temporal order of appearance in the values, and it is clear that, on some occasions at least, the order may be material. To take an extreme case, suppose we are told that in a sample of 100 births 53 are male. We conclude that the sample is concordant with the hypothesis that male and female births occur at random with probability $\frac{1}{2}$. But if we knew in addition that the *first* 53 births were male and the next 47 female we should almost certainly reject the hypothesis.

21.40. If sampling is conducted by taking members one at a time from a population and the process is random, then any order is as probable as any other order. The sample

may be considered as a section of an infinite series generated by the sampling process, and this series ought to behave like von Mises' Irregular Kollektiv (**7.15**). It is a happy hunting-ground for the theorist, since there is no limit to the number of tests which can be invented to ascertain whether a given finite series conforms to the random scheme. We have considered a few such tests in connection with random sampling numbers (**8.15**) and shall discuss others in connection with time-series (Chapter 30). Here we discuss a few tests which are useful in detecting departures from randomness in the sampling. We are not now considering hypotheses as to the parent population, but since the randomness of the sampling is an essential element of inferences in probability it is convenient to consider the reliability of the sampling, together with inferences from the sample about the parent.

Ranking Tests

21.41. Suppose we have a sample of n members $x_1 \ldots x_n$, in that order, and are doubtful about its randomness. Such doubts may arise owing either to defects in the sampling or to possible alterations in the population while the sampling is going on. In the first case the process itself is at fault; in the second, circumstances are at work to make the sample something other than it purports to be, a random sample from a single population. Either influence may relate the magnitude of the x's to the order in which they occur, and the values $x_1 \ldots x_n$ are not then a random order in the sense that any other order was equally probable.

Let us then consider all the possible orders, $n!$ in number, of the observed values $x_1 \ldots x_n$. A proportion of these, determined by a significance level of 5 per cent. or 1 per cent., say, we will decide to reject as improbable; and we will select as the "improbable" rankings those which exhibit the systematic appearance of which we are afraid, and particularly the regular rise or fall from x_1 to x_n in magnitude. In short, we rank the sample in order of magnitude, say $X_1 \ldots X_n$, where the X's are a permutation of the first n integers, and compute a rank correlation coefficient between this order and the order $1 \ldots n$. If the coefficient is large in absolute value ("large" being determined by the significance level) we suspect the sample of being subject to systematic influences.

Example 21.10

Thirty persons in the income group £1000–£1500 are asked to supply returns of their annual income for some purpose connected with taxation. It is intended to summarise their replies by a given date, but when that date arrives only 20 answers have been received. This is a frequent event in postal inquiries, even when the return is compulsory, and it has to be decided whether the 20 returns may be accepted as representative of the 30. There are prior reasons for suspecting that persons with bigger incomes may delay more than the others, partly because of difficulty in completing returns and partly because of a natural reluctance to part with information which may tell against them.*. We therefore wish to ascertain from the 20 returns whether there is any evidence that persons with smaller incomes tend to submit returns earlier than those with larger incomes.

Suppose the 20 returns give incomes, in that order, of £ per annum : 1180, 1270, 1400,

* This is an assumption for the purposes of the example and not intended as a statement about taxation returns in real life.

1090, 1190, 1250, 1170, 1300, 1290, 1310, 1280, 1350, 1320, 1380, 1420, 1390, 1470, 1360, 1220, 1460. The ranking order is—

No. of sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rank	3	7	17	1	4	6	2	10	9	11	8	13	12	15	18	16	20	14	5	19
Difference	− 2	− 5	− 14	3	1	0	5	− 2	0	− 1	3	− 1	1	− 1	− 3	0	− 3	4	14	1

The sum of squares of differences is 508 and thus the Spearman coefficient of rank correlation between observed and natural order $1 \ldots n$ is

$$\rho = 1 - \frac{6 \times 508}{7980} = 0{\cdot}618.$$

The probability of obtaining such a value or greater (**16.18**) may be found from "Student's" distribution by putting

$$t = \rho \left(\frac{n-2}{1-\rho^2} \right)^{\frac{1}{2}} = 3{\cdot}34$$

$$\nu = 18,$$

and is found from Appendix Table 3, vol. I, to be about 0·004. The test confirms our suspicion that size of income is correlated with order of appearance, and if we intend to use the mean income of the 20 returns as an estimate of the income in the full 30 we must recognise that it may very well be an under-estimate.

21.42. It will be noted in this example that we have made no assumption about the distribution of incomes in the sample or the population (the latter of which would certainly not be normal) and have used the sample values themselves without any reference to the question whether they were representative. This does not invalidate our inference, which is made within the population of samples obtained by permuting the observed values. (Cf. **17.44** and **17.45**.)

21.43. A second test of use in random series, particularly when it is suspected that cyclical effects are present, may be obtained by counting the occurrences of "peaks" or "troughs" in the series. A member is said to be a "peak" if it is greater than the two neighbouring members, and a "trough" if it is less than those members. In either case it is a "turning-point". The interval between turning-points is called a "phase".

Three consecutive observations are required to define a turning-point. If the series is random the probability that any given three provides a turning-point is $\frac{2}{3}$, for the values x_1, x_2, x_3 may occur in six orders and in only four is the greatest or least value the middle one. In a series of N terms there are $N - 2$ sets of three, and hence the expected number of turning-points p is

$$E(p) = \tfrac{2}{3}(N-2). \qquad (21.68)$$

The variance and higher moments of p are not so easy to determine. Like the ranking problems considered in Chapter 16 (to which the present problem is analogous), the distributions resulting are rather complicated. We quote without proof the results

$$\mu_2(p) = \frac{16N - 29}{90} \qquad (21.69)$$

$$\mu_3(p) = -\frac{16(N+1)}{945} \qquad (21.70)$$

$$\mu_4(p) = \frac{448N^2 - 1976N + 2301}{4725}. \qquad (21.71)$$

As N tends to infinity the distribution tends to normality fairly rapidly, and p may, for finite N, be taken as normally distributed about mean $\frac{2}{3}(N-2)$ with variance $\frac{16N-29}{90}$.

21.44. A further test may be derived from the distribution of phase lengths. The probability of a phase of length d in a series of $d+1$ terms is clearly $\frac{2}{(d+1)!}$, for only two of the possible permutations are favourable. In a series of length N there are $N-d-2$ possible phases of length d, for $d+3$ points are required to determine the phase. The probability of a phase d in $d+3$ terms is

$$\left\{\frac{1}{(d+1)!}-\frac{1}{(d+2)!}\right\}-\left\{\frac{1}{(d+2)!}-\frac{1}{(d+3)!}\right\}=\frac{d^2+3d+1}{(d+3)!} \quad . \quad (21.72)$$

and hence the number of phases of length d is

$$N!\,\frac{2(N-d-2)(d^2+3d+1)}{(d+3)!} \quad . \quad . \quad . \quad . \quad . \quad (21.73)$$

Now the number of possible phases is

$$N!\left\{\frac{2N-7}{3}+\frac{2}{N!}\right\} \quad . \quad . \quad . \quad . \quad . \quad (21.74)$$

for there is one fewer phase than turning-points, $\frac{2}{3}(N-2)$ in number, and the whole series may be a phase, which accounts for the factor $2/N!$ In practice this is negligible, and for the probability of a phase d in a series of N we then have (21.73) divided by (21.74), namely

$$\frac{6(d^2+3d+1)(N-d-2)}{(d+3)!\,(2N-7)} \quad . \quad . \quad . \quad . \quad . \quad (21.75)$$

The moments of this distribution are easily obtained to a very close approximation. For example,

$$\mu_1'(d)=\frac{6}{2N-7}\sum_{d=1}^{N-3} d\,.\,\frac{(N-d-2)(d^2+3d+1)}{(d+3)!}$$

$$=\frac{6}{2N-7}\sum_{1}^{N-3}[(N-2)\{(d+3)(d+2)(d+1)-3(d+3)(d+2)+5(d+3)-3\}$$
$$-(d+3)(d+2)(d+1)d+3(d+3)(d+2)(d+1)-8(d+3)(d+2)$$
$$+13(d+3)-9]/(d+3)!$$

$$=\frac{6}{2N-7}\Sigma\left[(N-2)\left\{\frac{1}{d!}-\frac{3}{(d+1)!}+\frac{5}{(d+2)!}-\frac{3}{(d+3)!}\right\}\right.$$
$$\left.-\frac{1}{(d-1)!}+\frac{3}{d!}-\frac{8}{(d+1)!}+\frac{13}{(d+2)!}-\frac{9}{(d+3)!}\right]$$

Remembering the rapid convergence of $\sum_0^N \frac{1}{x!}$ to e, we may write this as

$$\frac{6}{2N-7}[(N-2)\{e-1-3(e-2)+5(e-\tfrac{5}{2})-3(e-\tfrac{8}{3})\}$$
$$-e+3(e-1)-8(e-2)+13(e-\tfrac{5}{2})-9(e-\tfrac{8}{3})].$$

$$\mu_1'(d)=\frac{3(N+7-4e)}{2N-7}\sim\frac{3}{2}. \quad . \quad . \quad . \quad . \quad (21.76)$$

Similarly we find

$$\mu_2(d) = \frac{3}{(2N-7)^2}\{(8e-21)N^2 + (4e-17)N - (48e^2 - 140e + 14)\} \sim 0{\cdot}560. \quad (21.77)$$

21.45. In comparing observed distributions of phases with expected values the ordinary χ^2-test cannot be applied, because the probabilities of the events in a finite series are not independent. A test of significance has been derived by Wallis and Moore (1941), who consider a grouping into three categories, $d = 1$, $d = 2$ and $d \geqslant 3$. They conclude that χ^2 calculated from these three groups can be tested in the usual Type III form with $\nu = 2\frac{1}{2}$ if $\chi^2 \geqslant 6{\cdot}3$. For lower values $\frac{6}{7}\chi^2$ can be tested in that form with $\nu = 2$.

This test is independent of the law of distribution of the variables and is thus of general application. It has to be remembered, however, that generality in these matters may be offset by loss of sensitivity, and more searching tests may be required in certain cases.

Example 21.11

The following table shows the deviations from a moving nine-year average of potato yields in England and Wales for the years 1888–1935 (units are $\frac{1}{10}$th ton) :—

Year.	Yield.	Year.	Yield.	Year.	Yield.	Year.	Yield.
1888	− 6	1900	− 7 *T*	1912	− 15 *T*	1924	1 *T*
89	+ 2 *P*	01	+ 6 *P*	13	+ 3 *P*	25	+ 2 *P*
90	− 4 *T*	02	− 3	14	+ 2	26	− 9 *T*
91	− 3	03	− 7 *T*	15	+ 1	27	− 3
92	− 1	04	+ 2 *P*	16	− 2 *T*	28	+ 9 *P*
93	+ 6 *P*	05	0 *T*	17	+ 5 *P*	29	+ 5
94	− 2 *T*	06	+ 1 *P*	18	+ 4	30	+ 1
95	+ 7 *P*	07	− 7 *T*	19	− 4 *T*	31	− 10 *T*
96	+ 3	08	+ 8 *P*	20	− 3 *P*	32	+ 1
97	− 6 *T*	09	+ 4	21	− 9 *T*	33	+ 2
98	+ 2 *P*	10	+ 3 *T*	22	+ 11 *P*	34	+ 5 *P*
99	0	11	+ 4 *P*	23	− 1 }	35	4

We have marked with *P* and *T* the peaks and troughs of the series. The observed number of turning-points is 31 in a series of 48 terms. The expected number is, from (21.68), $\frac{2}{3}(48 - 2) = 30{\cdot}67$, almost exactly the number observed. No test of significance is required.

The duration of phases is :—

	Observed	Predicted (21.75)
$d = 1$	20	18·75
2	6	8·07
3 and over	4	3·18
	30	30·00

Here, again, a test is hardly necessary. We find, in fact, $\chi^2 = 0{\cdot}826$, $\frac{6}{7}$ of which for $\nu = 2$ is not significant.

We conclude that these tests provide no evidence against the randomness of the series and hence do not suggest any cyclical movement in the yields.

21.46. In the foregoing example we have treated the two values in 1923 and 1924 as a single value since they are equal. These so-called "ties" frequently occur in ranking work and are a great nuisance. In the present case there is only one, and any reasonable method of treating it will not affect the test. Where "ties" are numerous enough to make a serious difference some systematic method of treating them is desirable, particularly if more than two individuals are tied. They may be treated as a single observation, as in this case (although it would probably be better then to reduce N accordingly); or, preferably, they may be counted as a mean value, e.g. with a tied pair we should consider the first as greater than the second and then the second greater than the first, counting the number of turning-points or phases as one-half in each case and adding the two together. This, as in all similar ranking problems, makes the theoretical discussion of sampling very complicated, and if it is desired to make a precise use of significance tests a further possibility is to assume that the tied members are ranked in the order most unfavourable to the hypothesis under test, so as to be on the safe side.

Conditional Tests

21.47. When several unknown parameters are concerned, it may be difficult to find a sampling distribution dependent only on one of them which will form a basis for estimation or a test of significance. Sometimes, however, we can get rid of undesirable parameters by restricting the distribution in some way, and particularly by considering a distribution of samples which have some specified quality in common with the observed sample. Such distributions we shall, in Bartlett's phrase, call conditional. Fisher expresses a similar idea by speaking of samples which have the same *configuration.*

The most important application of this principle is in the testing of regression coefficients, which we shall consider in the next chapter. Here we give a simple illustration of the method for the Poisson distribution.

Example 21.12

Suppose we have two samples from populations which are known to give the Poisson type of distribution but may have different parameters. We wish to determine whether the populations could be identical.

Suppose the frequencies of successes in the two samples are r_1 and r_2. If λ is the parameter of the parent (assumed the same for each), the probabilities of the samples are

$$e^{-\lambda}\frac{\lambda^{r_1}}{r_1!} \qquad \text{and} \qquad e^{-\lambda}\frac{\lambda^{r_2}}{r_2!},$$

and their joint probability is accordingly

$$P\{r_1, r_2 \mid \lambda\} = \frac{e^{-2\lambda}\lambda^{r_1+r_2}}{r_1!\,r_2!}. \qquad (21.78)$$

This depends on λ and does not help us in answering the question. However, for the probability of a sample with $r_1 + r_2$ successes we have (since the sum of two Poisson variates with parameters λ_1, λ_2 is distributed in the same form with parameter $\lambda_1 + \lambda_2$) :—

$$P\{r_1 + r_2 \mid \lambda\} = \frac{e^{-2\lambda}(2\lambda)^{r_1+r_2}}{(r_1 + r_2)!},$$

and hence

$$\frac{P\{r_1, r_2 \mid \lambda\}}{P\{r_1 + r_2 \mid \lambda\}} = \frac{(r_1 + r_2)!}{2^{r_1+r_2}\, r_1!\, r_2!} = \frac{r!}{2^r\, r_1!\, r_2!} \quad . \quad . \quad . \quad (21.79)$$

where $r = r_1 + r_2$.

Now in accordance with Bayes' theorem we have

$$P\{r_1, r_2 \mid \lambda\} = P\{r_1, r_2 \mid r_1 + r_2\}\, P\{r_1 + r_2 \mid \lambda\}$$

and hence

$$P\{r_1, r_2 \mid r\} = \frac{r!}{2^r\, r_1!\, r_2!}. \quad . \quad . \quad . \quad . \quad (21.80)$$

Consequently, *if we confine our attention to samples for which the total number of successes is* r, the probability of the observed r_1 and r_2 is independent of λ and is, in fact, the corresponding term in the binomial $(\frac{1}{2} + \frac{1}{2})^r$. The probability is clearly that of a partition of r into the observed r_1 and r_2, and if it is small we suspect the hypothesis that the samples emanated from the same population.

This kind of conditional inference raises the same sort of point as we noticed in **17.44.** We decide beforehand that, whatever r turns out to be, we will make the inference in the population of samples which yield that value of r.

Pitman's Tests

21.48. In the extreme conditional case we may consider an inference in a population of samples the members of which are the same as those actually observed, the population being given by permutations or partitions of the observed values. The tests of ranking and periodicity given above are cases of this kind. A similar procedure has been advocated by Fisher in the analysis of variance and the design of experiments, and will be considered in due course. We now proceed to examine tests of the same nature proposed by Pitman (1937*a*, 1938).

Suppose we have two sets of values $u_1 \ldots u_m$ and $v_1 \ldots v_n$ with means $\bar{u}$ and $\bar{v}$ and the mean of the two together equal to $\bar{z}$. Given $m + n$ objects, there are $\binom{m+n}{n}$ ways, say N, of separating them into two sets of m and n objects, of which the given set is one. We call $|\bar{u} - \bar{v}|$ the *spread* of the separation. Since

$$m\bar{u} + n\bar{v} = (m + n)\,\bar{z},$$

we have also for the spread

$$\frac{(m + n)\,|\bar{u} - \bar{z}|}{n} = \frac{(m + n)\,|\Sigma(u) - m\bar{z}|}{mn} \quad . \quad . \quad . \quad (21.81)$$

Take a probability $1 - \alpha = M/N$, where M is an integer. If R is a particular separation, and the number of separations with spread not less than that of R is not greater than M, we call R *discordant*. If there are M or more with a greater spread we call it *concordant*. A separation which is neither concordant nor discordant is called *neutral*. If $m = n$ the separations occur in pairs with equal spreads, and we then take M to be even. The discordant separations are most easily picked out as those with the largest values of $|\Sigma u - m\bar{z}|$.

If the observed separation is arrived at by chance, the probability that it is discordant is $M/N = 1 - \alpha$ when there are no neutral separations. If such exist, the probability

is less than $1 - \alpha$. Similarly the probability that a separation is concordant is $1 - \alpha$, or more, as the case may be.

Two samples $u_1 \ldots u_m$ and $v_1 \ldots v_n$ are said to be discordant, concordant or neutral according as the separations u and v are so. Having selected our significance points dependent on α, and hence having fixed M, we can find for what values of the spreads a pair of samples is discordant or otherwise, and hence whether our observed pair is so. If they are discordant we reject the hypothesis that they came from the same population.

Example 21.13 (Pitman, 1937*a*)

Two samples have the following values :—

$$0,\ 11,\ 12,\ 20$$
$$16,\ 19,\ 22,\ 24,\ 29.$$

Are they significantly different ?

There are 9 members altogether and hence $\binom{9}{5} = 126$ separations into samples of five and four. We take α to be as near as possible to 0·95, corresponding to a 5-per-cent. level of significance, and hence $M = 6$. We then find the groups which have the largest values of the spread. We have $\bar{z} = 17$, so that $m\bar{z} = 68$, and using the form $|\Sigma u - 68|$ we find those groups of four from

$$0,\ 11,\ 12,\ 16,\ 19,\ 20,\ 22,\ 24,\ 29,$$

which give the maximum value to this quantity. They are—

	$\lvert \Sigma u - 68 \rvert$
0, 11, 12, 16	29
0, 11, 12, 19	26
0, 11, 12, 20	25
29, 24, 22, 20	27
29, 24, 22, 19	26
29, 24, 20, 19	24

The group 0, 11, 12, 20 gives the fifth largest spread, and so with $M = 6$ the observed separation is discordant. Our inference is that the samples come from different populations. Only in four other cases out of 126 should we get so large a spread in samples from the same population.

21.49. The extended use of the above test is barred by practical inconvenience, but an approximate form based on a different measure of discordance may be used. We now put

$$w = \frac{m(\bar{u} - \bar{z})^2}{(N - m)\mu_2}, \qquad \qquad (21.82)$$

where μ_2 is the variance of the samples taken together and is thus a constant. The function w is hence linear in $(\bar{u} - \bar{z})^2$, the device of squaring, as usual, getting rid of difficulties associated with the use of the modulus $|\bar{u} - \bar{z}|$. N here refers to the total sample $m + n$.

Now, for the moments of $\bar{u} - \bar{z}$ we may use the results of **11.26** (vol. I, p. 284), giving the moments of the mean in sampling from a finite population ; for $\bar{z}$ is the population

mean. Replacing n in the formulae of that section by m and putting $N = m + n$, we have—

$$E(\bar{u} - \bar{z}) = 0$$

$$E(\bar{u} - \bar{z})^2 = \frac{N - m}{(N-1)\,m}\mu_2$$

$$E(\bar{u} - \bar{z})^4 = \frac{(N-m)\;[\{N^2 + N - 6m(N-m)\}\mu_4 + 3N(N-m-1)(m-1)\mu_2^2]}{m^3(N-1)(N-2)(N-3)}$$

and hence for the first two moments of w we find

$$E(w) = \frac{1}{N-1} \qquad (21.83)$$

$$E(w^2) = \frac{3}{N^2 - 1}(1 + \theta), \qquad (21.84)$$

where

$$\theta = \frac{N+1}{3(N-2)(N-3)}\left\{\frac{N(N+1)}{m(N-m)} - 6\right\}\left\{\gamma_2 + \frac{6}{N+1}\right\}, \qquad (21.85)$$

γ_2 referring to the measure of kurtosis $\frac{\mu_4}{\mu_2^2} - 3$.

For fixed N the modulus of the second factor in (21.85) will be found to have a maximum at $\frac{2(N-2)}{N}$ when $m = \frac{1}{2}N$, and it takes this value again at

$$\frac{N - 2m}{N} = \pm\sqrt{\frac{N-2}{2N-1}},$$

giving $\frac{m}{N-m} = \frac{1}{5}$ or 5 for $N = 14$ and wider limits for larger N. It will also be found that for $N > 6$ the factor $\frac{N(N+1)}{m(N-m)} - 6$ is not greater in absolute value than $\frac{2(N-2)}{N}$ if

$$\frac{1}{5} < \frac{m}{N-m} < 5,$$

i.e. unless one sample is more than four times as big as the other. Thus for such values and γ_2 not large, θ is small, and approximately

$$E(w^2) = \frac{3}{N^2 - 1}. \qquad (21.86)$$

Similarly, using the fact that for large m and N

$$E(\bar{u} - \bar{z})^{2r} = 1.3.5 \ldots (2r-1)\left(1 - \frac{m}{N}\right)^r \frac{\mu_2^r}{m^r},$$

we find approximately

$$E(w^3) = \frac{3.5}{(N-1)(N+1)(N+3)}. \qquad (21.87)$$

The moments given by (21.83), (21.86) and (21.87) are those of the B-distribution

$$dF = \frac{1}{B(\frac{1}{2}, \frac{1}{2}N - 1)}(1 - w)^{\frac{1}{2}N - 2}\, w^{-\frac{1}{2}}\, dw, \qquad (21.88)$$

which can therefore be used to approximate to the distribution of w. In point of fact the distribution seems to be remarkably close.

w may also be written

$$w = \frac{\frac{mn}{m+n}(\bar{u}-\bar{v})^2}{\Sigma(u-\bar{u})^2 + \Sigma(v-\bar{v})^2 + \frac{mn}{m+n}(\bar{u}-\bar{v})^2}, \qquad (21.89)$$

which shows that $w \leqslant 1$.

We also have

$$\frac{w}{1-w} = \frac{\frac{mn}{m+n}(\bar{u}-\bar{v})^2}{\Sigma(u-\bar{u})^2 + \Sigma(v-\bar{v})^2} \qquad (21.90)$$

and it is instructive to observe that the function on the right is the same as that of $\frac{u^2}{n_1+n_2-2}$ of (21.32) with a few changes of notation. A transformation of (21.88) to "Student's" form will in fact show that we can test $\sqrt{\frac{w\nu}{1-w}}$ in the t-distribution with $\nu = m+n-2$; for (21.88) then becomes

$$dF \propto \frac{du}{\left(1 + \frac{u^2}{m+n-2}\right)^{\frac{1}{2}(m+n-1)}} \qquad (21.91)$$

where

$$u = \sqrt{\frac{w\nu}{1-w}}. \qquad (21.92)$$

21.50. A test of a similar kind may be evolved for the product-moment correlation. Suppose we have two samples $x_1 \ldots x_n$ and $y_1 \ldots y_n$ and calculate

$$r = \frac{\text{cov}\, xy}{\sqrt{(\text{var}\, x \;\text{var}\, y)}}$$

for every possible pairing of the x's and y's, $n!$ in number. As before, if we choose an α and hence a number M such that $1 - \alpha = M/n!$ we may determine those pairings for which r is greatest and reject the hypothesis that x and y are independent in such cases if they fall among the M greatest. Since the denominator of r is constant, this is equivalent to attributing significance to the values of $|\Sigma xy - n\bar{x}\bar{y}|$ which exceed a given value determined by α.

Taking $\bar{x} = \bar{y} = 0$, without loss of generality we find

$$E(r) = 0 \qquad (21.93)$$

$$E(r^2) = \frac{1}{n^2 \,\text{var}\, x \;\text{var}\, y} E(\Sigma xy)^2$$

$$= \frac{1}{n-1}, \qquad (21.94)$$

and similarly, if γ_1, γ_2 are the modified measures of skewness and kurtosis for x (expressed in terms of k-statistics, i.e. $\gamma_1 = \dfrac{k_3}{k_2^{3/2}}$, $\gamma_2 = \dfrac{k_4}{k_2^2}$) and γ_1' and γ_2' those for y, it will be found that

$$E(r^3) = \frac{n-2}{n(n-1)^2}\gamma_1\gamma_1' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21.95)$$

$$E(r^4) = \frac{3}{(n-1)(n+1)} + \frac{(n-2)(n-3)}{n(n+1)(n-1)^3}\gamma_2\gamma_2'. \quad . \quad . \quad (21.96)$$

Thus to order n^{-1} we have

$$\left.\begin{aligned} E(r) &= E(r^3) = 0 \\ E(r^2) &= \frac{1}{n-1} \\ E(r^4) &= \frac{3}{(n-1)(n+1)} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (21.97)$$

These are the first four moments of the distribution

$$dF = \frac{1}{B(\frac{1}{2}, \frac{1}{2}n - 1)}(1 - x^2)^{\frac{1}{2}n-2}\,dx, \qquad -1 \leqslant x \leqslant 1. \quad . \quad . \quad (21.98)$$

Thus r may be tested in this distribution or equivalently, putting

$$t = \frac{r}{\sqrt{(1-r^2)}}\sqrt{(n-2)}, \quad . \quad . \quad . \quad . \quad (21.99)$$

in "Student's" form with $\nu = n - 2$.

In particular, if the numbers x and y reduce to rankings, we have the test already introduced in **21.41**. Compare also the result given for the distribution of Spearman's ρ in **16.18** (vol. I, p. 401).

The Combination of Tests

21.51. It sometimes happens that we have a number of tests of significance, all yielding various probabilities, which we wish to express as a single probability. Suppose, for instance, that we conduct an experiment five times and that some test, such as that of the mean, gives probabilities to the observed deviations of 0·2, 0·8, 0·01, 0·1, 0·03. In the ordinary way two of these values would be regarded as significant and the other three not. What conclusion are we to draw as to the five taken together?

Suppose we have k values of the probability, $p_1 \ldots p_k$. The distribution of any particular p is rectangular, i.e.

$$dF = dp \qquad 0 \leqslant p \leqslant 1.$$

Hence, if $x = -\log p$ the distribution of x is

$$dF = e^{-x}\,dx, \qquad 0 \leqslant x \leqslant \infty$$

and its characteristic function is

$$\begin{aligned} \phi(t) &= \int_0^\infty e^{itx - x}\,dx \\ &= \frac{1}{1 - it}. \end{aligned}$$

Hence if we write

$$A = -\sum_{j=1}^{k} \log p_j, \quad . \quad . \quad . \quad . \quad . \quad (21.100)$$

the distribution of A has a characteristic function

$$\phi(t) = \frac{1}{(1-it)^k},$$

and is therefore given by

$$dF = \frac{1}{\Gamma(k)} A^{k-1} e^{-A}\, dA. \quad . \quad . \quad . \quad . \quad (21.101)$$

Putting

$$M^2 = 2A = -2\Sigma \log p = -2 \log \Pi p \quad . \quad . \quad . \quad (21.102)$$

we see that the distribution of M^2 is

$$dF \propto M^{2k-1} \exp(-\tfrac{1}{2}M^2)\, dM \quad . \quad . \quad . \quad . \quad (21.103)$$

or M^2 is distributed as χ^2 with $\nu = 2k$ degrees of freedom.

Example 21.14 (K. Pearson, 1933*b*, quoting data from E. M. Elderton, 1933).

Pairs of boys were selected in various age-groups and one member of each pair fed on raw, the other on pasteurised milk. The differences in gain in weight are shown in the following table, together with the standard errors of the differences based on large-sample theory.

(1) Age-group. (Central value in years).	(2) Number of Pairs.	(3) Mean Difference in Weight Gained, Raw less Pasteurised.	(4) Standard Error of Difference.	(5) Probability of Observed Difference or Greater, p_k.	(6) $\log_{10} p_k$.
$6\frac{3}{4}$	73	− 0·066	0·054	0·8888	$\bar{1}$·9488
$7\frac{3}{4}$	76	+ 0·022	0·053	0·3409	$\bar{1}$·5326
$8\frac{3}{4}$	71	− 0·003	0·052	0·5239	$\bar{1}$·7193
$9\frac{3}{4}$	77	+ 0·011	0·055	0·4207	$\bar{1}$·6240
$10\frac{3}{4}$	60	+ 0·002	0·057	0·4840	$\bar{1}$·6849
					$\bar{2}$·5096

The values of p_k in column (5) are obtained by expressing the observed deviations in column (3) in terms of the standard error in column (4) and hence determining the probability from the normal integral. We have

$$M^2 = -2\,\Sigma \log_e p = -2\,\frac{\Sigma \log_{10} p}{\log_{10} e}$$

$$= 6{\cdot}86$$

$$\nu = 10.$$

The probability of a value of $\chi^2 \geqslant 6{\cdot}86$ for $\nu = 10$ is about 0·74, and the test as a whole does not support the hypothesis of a differential effect on feeding between the two kinds of milk.

Nuisance Parameters

21.52. From the foregoing it will have been clear that in the theories of both estimation and significance one of the main problems is to find a distribution which is independent of certain unknown parameters in the parent population. Parameters of this kind, necessary as they are in the specification of the parent and the precise formulation of our problem, can be a nuisance when we are seeking to make exact statements about some other parameter on which interest is focussed. For this reason they have been named nuisance parameters. It may be useful if at this point we summarise the methods available for getting rid of them.

(*a*) First of all there is the process of " Studentisation ", whereby we can remove scale parameters from the sampling distribution by a suitable choice of statistic. (Cf. **19.26**.)

(*b*) Secondly, we may restrict the inference to a sub-population which is conditioned by having certain values in common with the observed sample. It sometimes happens that the distribution in this sub-population does not contain the nuisance parameters, whereas a distribution in the full population would do so (**21.47**).

(*c*) In the comparison of two samples, or even the testing of a single sample involving an unknown mean, that parameter may be eliminated by differencing (**21.27**). As regards the case of the single sample, it is clear that if $x_1 \ldots x_n$ are independent and n is even, the values $x_1 - x_2, x_3 - x_4, \ldots x_{n-1} - x_n$ will also be independent and be distributed with zero mean (though of course there are only $\frac{1}{2}n$ of them).

(*d*) Transformations of the variate may sometimes either eliminate the nuisance parameter altogether or reduce its importance. The most noteworthy case is Fisher's transformation of the correlation coefficient (**14.18**, vol. I, p. 345). The transformed function $z - \zeta$ is distributed nearly normally with variance $1/(n - 3)$, so that the difference of two correlations when transformed does not involve the common value of ζ. (Cf. Example 14.8.)

(*e*) We may find distributions which are independent of the unknown parameters, and even of the population, by using the methods of ranking or considering partitions (**21.41**, **21.48**).

(*f*) The fiducial argument, in at least one known case, gives a test independent of unknown parameters, namely the Behrens test (**20.13**).

It must be realised, however, that all these types of inference do not stand on equal footings. In particular (*e*) requires further examination, as we proceed to show.

21.53. We may now review the many different tests which have been described in this chapter and consider more closely the type of reasoning on which they are based. We may group our tests broadly into two classes, those which give a direct test of a given value of a parent parameter and those which do not.

The first class rests on a type of inference which we have discussed fully in connection with the problem of estimation. There is, in fact, only a difference in viewpoint, and little or none in essential ideas, between estimating a parameter by assigning a range to acceptable values (whether by confidence intervals or fiducial intervals) and ascertaining whether some prior value lies in that range. The significance of parameters in large samples, the test of the mean in normal samples by " Student's " distribution, the test of a correlation coefficient in normal samples, and others of the same kind relating to a specified parameter have the same logical foundation as the theory of confidence intervals or the theory of

fiducial intervals, whichever is preferred. *They all provide for the consideration of alternative values of the parameter.*

21.54. The second group of tests are not, on the face of it, concerned with the value of a parameter in a parent population, and some of them take no account of possible alternative hypotheses. Consider, for example, a test of normality or a test of randomness. The hypothesis is that the population is normal or the sampling is random, as the case may be, but this does not specify a parameter. What alternatives to normality or to randomness are we considering, if any ? We must have the existence of such alternatives in mind, however vaguely, for otherwise we should not be testing these particular hypotheses. But can we say what they are ? And if not, do our inferences remain valid ? When working with a probability α shall we still be right in a proportion α of the cases in the long run ?

21.55. The kind of argument we have used in all these cases is this: on the given hypothesis the observed sample *and all samples providing a greater value of the statistic being used for the test* have a small probability. Therefore we reject the hypothesis.

We may note at once that in rejecting the hypothesis we do so in favour of another hypothesis for which the observations are more probable. We may not express this thought explicitly, but it is there. The various statistics we use for testing normality, for instance b_1, can arise with greater probability from other populations which are skew or have a marked deviation from mesokurtosis; the fact is assumed as self-evident (as indeed it is) and hence, if the statistic is improbable for the normal case there will be non-normal cases of greater probability. We remark, nevertheless, that the actual probability α is calculated *on the normal hypothesis* and does not hold for the non-normal cases. Thus we can no longer assert that we are right in proportion α of the cases. We are therefore relying on a less definite principle of inference to the effect that we reject a hypothesis which gives an improbable value to observation, provided that there exists some other hypothesis which gives a more probable value.

21.56. A similar argument applies to tests of randomness. It is obvious that many other methods of generating a series exist which give a greater probability to a systematic series than the random method, and in rejecting the latter we do so more or less consciously in favour of the former. Our intuitive feelings on the point lead us to apply one test when we have the possibility of systematic order in mind (the ranking test) and another when we are interested in oscillations (the phase test). What we are doing, in effect, is selecting the test of randomness which we feel to discriminate best between the hypothesis of randomness and the alternative possibilities.

21.57. Although, therefore, much remains to be done in putting tests of normality, randomness and goodness of fit on a formal logical basis, there do not appear to be any serious difficulties in doing so insofar as the specification of alternative hypotheses is concerned. But there remains the difficulty hinted at at the beginning of 21.55. In the majority of cases we have a probability $1 - \alpha$ that the observed statistic t_0 will be exceeded, and if this is small reject the hypothesis. But why *exceeded* ? Why reject the hypothesis because of the improbability of a number of events which have not happened ?

Here also it seems that a closer inquiry into the logic of the process would be worth while. We have seen how it can be justified by confidence-interval or fiducial theory

when a parameter is under consideration. When no parameter is specified, the process must, in the present state of our knowledge, rest on more intuitive ideas. My own view is that, in a vague kind of way, we are really considering the range of values of a parameter without realising it. In selecting a statistic to carry out the test, we usually relate it to the sort of effect we are expecting to divert the real state of affairs from those of our hypothesis. For instance, if we suspect cyclical effects in a random series we base a test on oscillations in that series. The further the series deviates from randomness the greater will be the value of our statistic; and consequently, if we could measure deviation from randomness (in the direction of cyclicality), we should have a parameter which could be located in a range in the manner of confidence intervals. Such a range would exclude the larger values of our statistic if it can be regarded in any sense as estimating the parameter (or, more generally, as increasing with it); and hence the procedure of rejecting the hypothesis if the statistic is among these large values may be justified.

21.58. It is for this reason that we began the chapter by defining tests of significance in relation to a parameter-value given *a priori*. It seems probable that in the ultimate analysis no other definition will be satisfactory. The fact that in this chapter we have given tests of hypotheses which do not appear to specify a parameter value is, I think, merely a reflection of the fact that the nature of those hypotheses and the inferences about them are not usually understood clearly but are based on more or less intuitive ideas. It is probable that many of these ideas are sound and can be given explicit logical foundation; but the matter awaits investigation by the statistical logician.

21.59. There remains for consideration the type of inference used in Pitman's tests (**21.48** and **21.49**). These are of the character of tests of randomness. Given a set of values, we consider all the arrangements in which they could have happened and reject the hypothesis if the observed arrangement is improbable. Here again, as it seems to me, there is a suppressed series of alternative hypotheses which would make the observed value more probable; and in choosing the test, such as the "spread" or the high value of a correlation, we are intuitively relating the magnitude of a statistic to the deviation from randomness. Pitman himself has shown, however, that when the hypothesis is definite and specifies the difference of two means, the tests give confidence intervals in the ordinary way (cf. Exercise 21.15.)

We shall resume the general theory of tests of significance in Chapter 26.

NOTES AND REFERENCES

For the use of the t-distribution in non-normal cases see Geary (1936*b*) and Bartlett (1935*a*), the latter of whom shows that, for moderate samples, departures from mesokurtosis are not very serious. For approximations to t in the normal case see Hendricks (1936) and Hotelling and Frankel (1938). For approximations to the z-distribution see Cochran (1940*a*), Cornish and Fisher (1937), and Paulson (1942). See also references to Chapter 23.

For the further theory of the χ^2-test see Neyman and Pearson (1928, 1931*a*) and for another test of goodness of fit Neyman (1937*a*). The theory of **21.44** has been studied by a number of writers, notably by André (1884), Kermack and McKendrick (1936, 1937), and Wallis and Moore (1941).

The amalgamation of tests given in **21.51** was apparently first given by Fisher in an

early edition of *Statistical Methods for Research Workers* and was studied in detail by K. Pearson (1933*b*) under the title of the P_λ-test, and by E. S. Pearson (1938).

For a test of significance of the difference of two variances in samples from a bivariate normal population see Hirschfeld (1937), Finney (1938), Pitman (1939*c*), Morgan (1939), and De Lury (1938); and see Exercise 21.3.

For the tests by Pitman, see his papers of 1937*a*, 1938. The similar problem in the testing of homogeneity in the analysis of variance has also been studied—see references to Chapters 23 and 24.

For the test of difference of means when variances are unequal from the point of view of confidence intervals see Welch (1938*b*) and the appendix to this paper by Miss Tanburn.

EXERCISES

21.1. For the population represented approximately by

$$dF = \frac{1}{\sqrt{(2\pi)}}\left\{1 - \frac{\kappa_3}{6}(3x - x^3)\right\} e^{-\frac{x^2}{2}} dx,$$

show that, if κ_3^2 is negligible, the joint probability of a sample $x_1 \ldots x_n$ differs from that if κ_3 is zero by a term

$$\frac{1}{(2\pi)^{\frac{n}{2}}} \frac{\kappa_3}{6}\left\{\sum_{j=1}^{n}(x_j^3) - 3\sum_{j=1}^{n}(x_j)\right\} \exp\left(-\tfrac{1}{2}\Sigma x_j^2\right) dx_1 \ldots dx_n.$$

By the transformation

$$y_1 = \frac{1}{\sqrt{2}}(x_1 - x_2)$$

$$y_2 = \frac{1}{\sqrt{6}}(x_1 + x_2 - 2x_3)$$

$$y_n = \frac{1}{\sqrt{n}}(x_1 + x_2 \ldots + x_n)$$

and the further transformation

$$\begin{aligned} y_1 &= \rho \sin\phi_{n-3} \sin\phi_{n-4} \ldots \sin\phi_1 \sin\phi_0 \\ y_2 &= \rho \sin\phi_{n-3} \sin\phi_{n-4} \ldots \sin\phi_1 \cos\phi_0 \\ y_3 &= \rho \sin\phi_{n-3} \sin\phi_{n-4} \ldots \cos\phi_1 \\ y_{n-1} &= \rho \cos\phi_{n-3}, \end{aligned}$$

show that the corrective term to the distribution of "Student's" t is

$$dt \int_0^\infty \left(\frac{1}{\nu^2}t^3\rho^3 + \frac{3}{\nu}t\rho^3 - \frac{3n}{\nu}t\rho\right) \exp\left\{-\frac{\rho^2}{2}\left(1 + \frac{t^2}{\nu}\right)\right\} \rho^\nu \, d\rho$$

and hence obtain equation (21.11).

(Geary, 1936*b*.)

21.2. By the polar transformation of the type of the previous exercise applied to all n variates show that if a random sample is drawn from a normal population with zero mean the frequency element may be written as

$$\frac{1}{(2\pi)^{\frac{n}{2}}}\rho^{n-1} e^{-\frac{1}{2}\rho^2} d\rho \, d\phi_0 \sin\phi_1 \, d\phi_1 \sin^2\phi_2 \, d\phi_2 \ldots \sin^{n-2}\phi_{n-2} \, d\phi_{n-2}.$$

Hence if $w = \dfrac{\Sigma |x|}{ns}$, where s^2 is the sample variance, the distribution of w is independent of that of s. Hence show that for the distribution of w, writing $a = \sqrt{\dfrac{2}{\pi}}$,

$$\mu_1' = \frac{\{\Gamma(\frac{1}{2}n+1)\}^2}{\Gamma(n+1)} \frac{2^n}{\sqrt{n}} a^2$$

$$\mu_2 = \frac{1}{n^2}\{n^{(1)} + a^2 n^{(2)}\}$$

$$\mu_3 = \frac{\mu_1' a}{n^2}\{2n^{(1)} + 3n^{(2)} + a^2 n^{(3)}\} \Big/ \frac{n+1}{n}$$

$$\mu_4 = \frac{1}{n^4}\{3n^{(1)} + (8a^2+3)n^{(2)} + 6a^2 n^{(3)} + a^4 n^{(4)}\} \Big/ \frac{n+2}{n}.$$

Hence show that for $n = 50$, $\sqrt{\beta_1} = -0{\cdot}24$ and $\beta_2 = 3{\cdot}10$, indicating fairly rapid tendency to normality.

(Geary, 1935*a*).

21.3. Show that in samples from a normal bivariate population

$$dF \propto \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right\}\right] dx\, dy,$$

the functions
$$u_j = \frac{x_j}{\sigma_1} + \frac{y_j}{\sigma_2}, \quad v_j = \frac{x_j}{\sigma_1} - \frac{y_j}{\sigma_2}$$

are distributed independently and that their correlation coefficient R may be written

$$R = \frac{a-\alpha}{\sqrt{\{(a+\alpha)^2 - 4a\alpha r^2\}}},$$

where
$$\alpha = \frac{\sigma_1^2}{\sigma_2^2}, \quad a = \frac{\Sigma(x-\bar{x})^2}{\Sigma(y-\bar{y})^2},$$

and r is the correlation between the observed x's and y's. Hence show that

$$t = \frac{R\sqrt{(n-2)}}{\sqrt{(1-R^2)}} = \frac{(a-\alpha)\sqrt{(n-2)}}{\sqrt{\{4(1-r^2)a\alpha\}}}$$

is distributed as "Student's" t with $n-2$ degrees of freedom. Show how to test the ratio α from this result.

(Pitman, 1939*c*. The test has the remarkable property of being independent of the parent correlation ρ.)

21.4. If an even number n of members of a sample come from a population with mean μ, show how to find a sample of half the size distributed with twice the variance about zero mean. Hence show how to extend the result of Exercise 21.2 to the case where the population mean is not zero.

21.5. If a parameter admits of a sufficient estimator, show that a test of its significance can be derived direct from the likelihood function.

21.6. Derive equations (21.47) and (21.48).

21.7. Let $l_{11}, l_{12} \ldots l_{1,n-1}$ be $(n-1)$ linear functions of the observations which are orthogonal to one another and to $\bar{x}_1$, and let them have zero mean and variance σ_1^2. Similarly define $l_{21} \ldots l_{1,n-2}$.

Then, in two samples of n from normal populations with equal means and variances σ_1^2 and σ_2^2, the function

$$\frac{\sqrt{n}\,(\bar{x}_1 - \bar{x}_2)}{\{\Sigma\,(l_{1j} + l_{2j})^2/(n-1)\}^{\frac{1}{2}}}$$

will be distributed as "Student's" t with $n-1$ degrees of freedom.

(Bartlett, 1937*c*, and Welch, 1938*b*. The test does not depend on the ratio σ_1^2/σ_2^2 and can be extended to the case of unequal sample numbers, but only at the expense of losing efficiency in the sense that the degrees of freedom number one less than the lower of the sample numbers.)

21.8. Given two samples of n_1, n_2 members from normal populations with unequal variances, show that by picking n_1 members at random from the n_2 (where $n_2 \geqslant n_1$) and pairing them at random with the members of the first sample, a test of significance of difference of means can be based on "Student's" distribution independently of the variance ratio in the populations. (This test, again, is exact, but sacrifices the information of $n_2 - n_1$ members of the second sample.)

21.9. If z is the ratio of the sample mean to sample standard deviation in normal samples, and n is large enough for the distribution of the variance to be regarded as normal, show that

$$c_n \sqrt{(2n)} \frac{z}{\sqrt{(z^2+2)}} = c_n \sqrt{(2n)} \frac{t}{\sqrt{\{t^2 + 2\,(n-1)\}}}$$

is distributed approximately normally with zero mean and unit variance, where

$$c_n = \sqrt{\frac{2}{n}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sim 1 - \frac{3}{4n} - \frac{7}{32n^2}.$$

(Hendricks, 1936.)

21.10. If x, y have a continuous frequency function $f(x, y)$, their characteristic function is

$$\phi(u, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(iux + ivy) f(x, y)\, dx\, dy.$$

Show that the distribution of x when y is given has a characteristic function

$$\phi(u \mid y) = \frac{\int_{-\infty}^{\infty} e^{-iyv}\,\phi(u, v)\, dv}{\int_{-\infty}^{\infty} e^{-iyv}\,\phi(0, v)\, dv}$$

(Bartlett, 1938*b*.)

21.11. If a set of parameters $\theta_1 \ldots \theta_p$ admit of a set of sufficient estimators, show that conditional inferences independent of $\theta_1 \ldots \theta_p$ are possible, the conditions being

22.3. We may also consider the more general curves typified by

$$Y = \frac{\int_{-\infty}^{\infty} y^r f(X, y)\, dy}{\int_{-\infty}^{\infty} f(X, y)\, dy}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.4)$$

the regression now being of the rth moment of y on x. If $r = 1$ we have the regression of the first moment, or simply the regression. If $r = 2$ and y is measured from the mean we have the so-called *scedastic* curve of y on x,

$$Y = \frac{\int_{-\infty}^{\infty} (y - \bar{y}_x)^2 f(X, y)\, dy}{\int_{-\infty}^{\infty} f(X, y)\, dy}, \quad . \quad . \quad . \quad . \quad . \quad (22.5)$$

which shows how the variance of y varies with x. Other forms which have been studied are the *clitic* curve

$$Y = \frac{\int_{-\infty}^{\infty} (y - \bar{y}_x)^3 f(X, y)\, dy}{\int_{-\infty}^{\infty} f(X, y)\, dy} \quad . \quad . \quad . \quad . \quad . \quad (22.6)$$

and the *kurtic* curve

$$Y = \frac{\int_{-\infty}^{\infty} (y - \bar{y}_x)^4 f(X, y)\, dy}{\int_{-\infty}^{\infty} f(X, y)\, dy} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.7)$$

These curves correspond to the moments of a univariate distribution, and the main characteristics of a bivariate form may be studied with their aid in much the same way as the lower moments can be used to summarise the properties of a univariate form.

22.4. It is interesting to remark that, just as we can find the moments direct from the characteristic function, so also we may ascertain the regressions of moments from the bivariate characteristic function, even when the distribution function itself is not explicitly given.

Let us write the frequency function in the form

$$f(x, y) = g(x)\, g_x(y), \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.8)$$

where $g(x)$ is the total frequency for any given x and $g_x(y)$ is the frequency of y for any given x. In the notation of the theory of probability we should write this

$$f(x, y) = g(x)\, g(y \mid x).$$

The characteristic function of x and y is then

$$\begin{aligned} \phi(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\{it_1 x + it_2 y\}\, g(x)\, g_x(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} e^{it_1 x}\, g(x)\, \phi_x(t_2)\, dx \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.9) \end{aligned}$$

where

$$\phi_x(t_2) = \int_{-\infty}^{\infty} e^{it_2 y}\, g_x(y)\, dy \quad . \quad . \quad . \quad . \quad . \quad (22.10)$$

and is the c.f. of y for a given x.

If the rth moment of y about the origin for a given x is μ'_{rx}, we have

$$i^r \mu'_{rx} = \left[\frac{\partial^r}{\partial t_2^r}\phi_x(t_2)\right]_{t_2=0}$$

and hence, from (22.9),

$$\left[\frac{\partial^r}{\partial t_2^r}\phi(t_1, t_2)\right]_{t_2=0} = i^r \int_{-\infty}^{\infty} e^{it_1 x} g(x)\, \mu'_{rx}\, dx \qquad (22.11)$$

Thus, by the Inversion Theorem,

$$g(x)\,\mu'_{rx} = \frac{(-i)^r}{2\pi}\int_{-\infty}^{\infty} e^{-it_1 x}\left[\frac{\partial^r}{\partial t_2^r}\phi(t_1, t_2)\right]_{t_2=0} dt_1, \qquad (22.12)$$

subject, of course, to conditions of existence. This gives us the required expression for μ'_{rx} in terms of x, and the regression can be written down at once.

22.5. Since

$$\phi(t_1, t_2) = \exp \sum_{j,k=0}^{\infty}\left\{\kappa_{jk}\frac{(it_1)^j}{j!}\frac{(it_2)^k}{k!}\right\}$$

we have

$$\left[\frac{\partial\phi}{\partial t_2}\right]_{t_2=0} = i \exp\left[\sum_{j=0}^{\infty}\left\{\kappa_{j0}\frac{(it_1)^j}{j!}\right\}\right]\sum_{j=0}^{\infty}\kappa_{j1}\frac{(it_1)^j}{j!}$$

$$= i\,\phi(t_1, 0)\sum_{j=0}^{\infty}\kappa_{j1}\frac{(it_1)^j}{j!} \qquad (22.13)$$

and $\phi(t_1, 0)$ may be written $\phi(t_1)$, being the characteristic function of $g(x)$. We also have, subject to existence conditions,

$$D^j g = \frac{d^j}{dx^j} g(x) = \frac{(-i)^j}{2\pi}\int_{-\infty}^{\infty} t_1^j e^{-it_1 x}\phi(t_1)\, dt_1. \qquad (22.14)$$

Hence, from (22.12), (22.13) and (22.14) we find

$$g(x)\,\mu'_{1x} = \frac{-i}{2\pi}\int_{-\infty}^{\infty} e^{-it_1 x}\left[\frac{\partial}{\partial t_2}\phi(t_1, t_2)\right]_{t_2=0} dt_1$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it_1 x}\phi(t_1)\sum_0^{\infty}\left\{\kappa_{j1}\frac{(it_1)^j}{j!}\right\} dt_1$$

$$= \sum_{j=0}^{\infty}\left\{\frac{\kappa_{j1}}{j!}(-D)^j g(x)\right\}, \qquad (22.15)$$

provided that the interchange of summation and integration in the last step is legitimate. Thus we have, for the regression of the mean,

$$Y = \sum_0^{\infty}\frac{\kappa_{j1}}{j!}\left[\frac{(-D)^j g(x)}{g(x)}\right]_{x=X} \qquad (22.16)$$

This notable result is due to Wicksell (1934*b*). The expansion is valid if the cumulants exist and if $g(x)$ and its derivatives are continuous in the range and zero at its extremes; for then the interchange of summation and integration in arriving at (22.15) is legitimate.

In particular, if $g(x)$ is normal and in standard measure we have

$$Y = \Sigma \frac{\kappa_{j1}}{j!} H_j(X), \qquad (22.17)$$

where $H_j(x)$ is the Tchebycheff-Hermite polynomial of order j (**6.20**, vol. I, p. 145).

Example 22.1

For the bivariate normal distribution about the mean we have

$$dF = k \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1 \sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right\} dx\, dy,$$

$$\phi(t_1, t_2) = \exp\{-\tfrac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)\}.$$

Hence

$$\left[\frac{\partial\phi}{\partial t_2}\right]_{t_2=0} = -\rho\sigma_1\sigma_2 t_1 \exp(-\tfrac{1}{2}\sigma_1^2 t_1^2),$$

and from (22.12)

$$g(x)\mu'_{1x} = \frac{i}{2\pi}\int_{-\infty}^{\infty} \rho\sigma_1\sigma_2 t_1 \exp\{-\tfrac{1}{2}\sigma_1^2 t_1^2 - it_1 x\}\, dt_1$$

$$= \frac{\rho\sigma_2}{\sigma_1^2\sqrt{(2\pi)}}\, x\, e^{-\frac{x^2}{2\sigma_1^2}}.$$

Hence

$$\mu_{1x} = \frac{\rho\sigma_2}{\sigma_1} x$$

and

$$Y = \frac{\rho\sigma_2}{\sigma_1} X,$$

the familiar relation of linearity for the regression of the mean of the normal distribution.

Alternatively, direct from (22.17) we have, since $\kappa_{j1} = 0,\ j > 1$

$$\frac{Y}{\sigma_2} = \kappa_{01} + \frac{\kappa_{11}}{\sigma_1} H_1(X)$$

$$Y = \frac{\rho\sigma_2}{\sigma_1} X, \text{ as before.}$$

Example 22.2 (Wicksell, 1934*b*)

Consider the frequency distribution of $\xi = \frac{1}{2}\Sigma(x^2)$ and $\eta = \frac{1}{2}\Sigma(y^2)$ where x, y are samples of n from the bivariate normal population

$$dF \propto \exp - \frac{1}{2(1-\rho^2)}\{x^2 - 2\rho xy + y^2\}\, dx\, dy.$$

The characteristic function is

$$\phi \propto \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(\tfrac{1}{2}x^2\theta_1 + \tfrac{1}{2}y^2\theta_2)\, dF\right]^n = \left\{(1-\theta_1)(1-\theta_2) - \rho^2\theta_1\theta_2\right\}^{-\frac{n}{2}},$$

where $\theta_1 = it_1$ and $\theta_2 = it_2$.

The distribution function cannot be expressed in a simple form, but we may determine the regressions without it. We have

$$\left[\frac{\partial^r\phi}{\partial\theta_2^r}\right]_{\theta_2=0} = \left(\frac{n}{2} + r - 1\right)^{[r]} \frac{\{1 - (1-\rho^2)\theta_1\}^r}{(1-\theta_1)^{\frac{1}{2}n+r}}.$$

Thus, from (22.12)

$$g(\xi)\mu'_{r\xi} = \frac{(-1)^r}{2\pi}\int_{-\infty}^{\infty} \frac{\left(\frac{n}{2} + r - 1\right)^{[r]} e^{-\theta_1\xi}\{1 - (1-\rho^2)\theta_1\}^r}{(1-\theta_1)^{\frac{1}{2}n+r}}\, d\theta_1.$$

The integrals may be evaluated by successive application of

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-\theta\xi}\,d\theta}{(1-\theta)^k} = \frac{1}{\Gamma(k)}\,\xi^{k-1}\,e^{-\xi},$$

and we find, for the regression of η on ξ,

$$\mu_{1\xi} = \frac{n}{2} + \rho^2\left(\xi - \frac{n}{2}\right)$$

$$\begin{aligned}\mu_{2\xi} &= \mu'_{2\xi} - (\mu'_{1\xi})^2 \\ &= (1-\rho^2)\left\{\frac{n}{2}(1-\rho^2) + 2\rho^2\,\xi\right\}.\end{aligned}$$

Thus the regressions of both mean and variance of η on ξ are linear.

Fitting of Curvilinear Regression Lines

22.6. From the practical point of view the case we have just considered, namely, the one where the distribution or characteristic function is given, is exceptional. The determination of regression curves has, in the majority of cases, to be carried out from numerically specified material, which we shall consider in the remainder of the chapter. We shall confine our attention to the regression of the mean.

In general the means of arrays will not lie exactly on a smooth curve (unless of course we choose a curve of order equal to the number of points to be fitted, less one). Nor do we know *a priori* what is the appropriate degree of a polynomial which will approximately represent the regression line. Let us, however, assume that the regression can be represented by a polynomial of order p:

$$Y = a_0 + a_1 X + a_2 X^2 + \ldots + a_p X^p. \qquad (22.18)$$

We will consider later how the appropriate value of p is to be determined in particular cases. Our problem is to determine the coefficients a from the data. As usual, we appeal to the principle of least squares, that is to say, we find the values of the a's which will minimise

$$U = \Sigma\,(y - a_0 - a_1 x - \ldots - a_p x^p)^2, \qquad (22.19)$$

the summation extending over the sample values.

Differentiating with respect to a_j, we have

$$\Sigma\,(x^j y) - a_0\,\Sigma\,x^j - a_1\,\Sigma\,x^{j+1} - \ldots - a_p\,\Sigma\,x^{j+p} = 0,$$

and similar equations for $j = 0, \ldots p$. Writing the moments without primes for simplicity and letting μ_j represent the jth moment of x, and μ_{j1} the bivariate moment $\Sigma\,(x^j y)$, we have

$$\left.\begin{array}{l} a_0\,\mu_0 + a_1\,\mu_1 + \ldots + a_p\,\mu_p = \mu_{01} \\ a_0\,\mu_1 + a_1\,\mu_2 + \ldots + a_p\,\mu_{p+1} = \mu_{11} \\ \quad\cdot \qquad \cdot \qquad \cdot \qquad \cdot \\ a_0\,\mu_p + a_1\,\mu_{p+1} + \ldots + a_p\,\mu_{2p} = \mu_{p1} \end{array}\right\} \qquad (22.20)$$

Writing now

$$\Delta^{(p)} = \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_p \\ \mu_1 & \mu_2 & \ldots & \mu_{p+1} \\ \cdot & \cdot & \ldots & \cdot \\ \mu_p & \mu_{p+1} & \ldots & \mu_{2p} \end{vmatrix} \qquad (22.21)$$

and $\Delta_j^{(p)}$ for the determinant obtained by substituting the product-moments $\mu_{01}, \ldots \mu_{p1}$ for the $(j + 1)$th column, we have, as the solution of (22.20),

$$a_j = \frac{\Delta_j^{(p)}}{\Delta^{(p)}}. \qquad (22.22)$$

22.7. It might appear that this solution could break down if $\Delta^{(p)} = 0$. Such a thing is not possible, however, except in the most trivial case. In fact, if the distribution function of the x's is $G(x)$, we have for $\Delta^{(p)}$

$$\Delta^{(p)} = \int\int \ldots \int \begin{vmatrix} 1 & x_0 & x_0^2 & \ldots & x_0^p \\ x_1 & x_1^2 & x_1^3 & \ldots & x_1^{p+1} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ x_p^p & x_p^{p+1} & x_p^{p+2} & \ldots & x_p^{2p} \end{vmatrix} dG_0\, dG_1 \ldots dG_p$$

or, if

$$D = \begin{vmatrix} 1 & x_0 & \ldots & x_0^p \\ 1 & x_1 & \ldots & x_1^p \\ \cdot & \cdot & \ldots & \cdot \\ 1 & x_p & \ldots & x_p^p \end{vmatrix}$$

$$\Delta^{(p)} = \int\int \ldots \int x_0^0\, x_1^1\, x_2^2 \ldots x_p^p\, D\, dG_0\, dG_1 \ldots dG_p.$$

If we now permute the suffixes of the x's in all possible ways and sum the $(p + 1)!$ resultants we obtain, in virtue of the definition of a determinant,

$$(p + 1)!\, \Delta^{(p)} = \int\int \ldots \int D^2\, dG_0\, dG_1 \ldots dG_p, \qquad (22.23)$$

and hence $\Delta^{(p)}$ is essentially positive.

22.8. From (22.18) and (22.22) we see that the regression line may be written

$$\begin{vmatrix} Y & 1 & X & \ldots & X^p \\ \mu_{01} & \mu_0 & \mu_1 & \ldots & \mu_p \\ \mu_{11} & \mu_1 & \mu_2 & \ldots & \mu_{p+1} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \mu_{p1} & \mu_p & \mu_{p+1} & \ldots & \mu_{2p} \end{vmatrix} = 0 \qquad (22.24)$$

This is a formal solution of our problem. The moments μ can be obtained from observation, and equation (22.24) then gives the regression line.

It will be observed that in order to preserve the symmetry we have written μ_0 for the total frequency unity.

22.9. A somewhat different approach leads to the same solution. If we assume that the regression line is a parabolic curve of order p, we may find the coefficients by the principle of moments. This would lead us to identify the lower moments

$$\Sigma (x^j y) = \Sigma x^j (a_0 + a_1 x + \ldots + a_p x^p)$$

as far as was necessary to determine the a's. This clearly leads back to equation (22.20).

Orthogonal Polynomials

22.10. The use of equation (22.24) in practice is subject to one serious drawback. If we have a set of data and no guide, apart from inspection, to the appropriate value of

p, the only course is to fit curves of order 1, 2, 3, . . . and so forth, until we reach the point when further terms do not improve the fit. Every time we add a new term the determinantal arithmetic has to be done afresh. To obviate this nuisance we shall consider the regression line in the form

$$Y = b_0 P_0 + b_1 P_1 + \ldots + b_p P_p, \quad . \quad . \quad . \quad . \quad (22.25)$$

where the P's are polynomials in X, P_j being of degree j. We shall determine the P's so that

$$\Sigma (P_j P_k) = 0, \qquad j \neq k \quad . \quad . \quad . \quad . \quad (22.26)$$

the summation extending over the observed values.

In minimising

$$\Sigma (y - b_0 P_0 - b_1 P_1 \ldots - b_p P_p)^2,$$

we shall have equations such as

$$\Sigma (yP_j) - b_0 \Sigma (P_0 P_j) - \ldots - b_p \Sigma (P_p P_j) = 0,$$

and in virtue of the orthogonal relations (22.26), this reduces to

$$\Sigma (yP_j) - b_j \Sigma (P_j^2) = 0. \quad . \quad . \quad . \quad . \quad (22.27)$$

Thus b_j is determined simply by P_j; and if, having fitted a curve of order p, we wish to go a step farther and add a term $b_{p+1} P_{p+1}$, the coefficients $b_0 \ldots b_p$ found from (22.27) remain unaltered.

22.11. Furthermore, the use of these orthogonal polynomials will give us a very convenient method of determining step by step the goodness of fit of the regression line. We have

$$\begin{aligned} U &= \Sigma (y - b_0 P_0 - \ldots - b_p P_p)^2 \\ &= \Sigma (y^2) - 2b_0 \Sigma (yP_0) - \ldots - 2b_p \Sigma (yP_p) + b_0^2 \Sigma (P_0^2) + \ldots + b_p^2 \Sigma (P_p^2). \end{aligned}$$

But from (22.27) we may express $\Sigma (yP_j)$ in terms of $\Sigma (P_j^2)$, and we thus find

$$U = \Sigma (y^2) - b_0^2 \Sigma (P_0^2) - \ldots - b_p^2 \Sigma (P_p^2). \quad . \quad . \quad . \quad (22.28)$$

Thus the effect of any term $b_j P_j$ is to reduce U by $b_j^2 \Sigma (P_j^2)$ and we may examine the effect of this term on U separately. If we find that the addition of any term $b_p P_p$ does not reduce U significantly, we may conclude that it is redundant (so far as concerns the representation of a regression line by a polynomial).

22.12. We proceed then to derive expressions for the orthogonal polynomials in the general case. Later we shall examine the important special case when the values of x are equidistant (as, for instance, with grouped data and most time-series).

Put

$$P_p = \sum_{j=0}^{p} c_{pj} X^j. \quad . \quad . \quad . \quad . \quad . \quad (22.29)$$

In this expression there are $(p + 1)$ unknown constants c, and hence in all the polynomials up to and including those of the pth order there are $\frac{1}{2} (p + 1) (p + 2)$ constants. The orthogonal relations up to and including order p will then provide $\frac{1}{2}p (p + 1)$ conditions

on the c's, so that $p + 1$ constants are assignable at will. We will take one for each P and assign it so that the coefficient of X^j in P_j is unity:

$$c_{jj} = 1. \qquad (22.30)$$

In particular $c_{00} = P_0 = 1$. The orthogonal relations are then just sufficient to determine the other c's. For instance, for the set c_{pj}, $j = 0 \ldots p-1$, they are

$$\begin{aligned} \Sigma P_p P_0 &= \Sigma P_p = 0 \\ \Sigma P_p P_1 &= 0 \end{aligned}$$

and so on. This system is clearly equivalent to the p equations

$$\left.\begin{aligned} \Sigma P_p &= 0 \\ \Sigma x P_p &= 0 \\ \cdot \quad \cdot \quad & \cdot \\ \Sigma x^{p-1} P_p &= 0 \end{aligned}\right\} \qquad (22.31)$$

On substituting for the P's from (22.29) we get

$$\begin{aligned} c_{p0}\mu_0 + c_{p1}\mu_1 + \ldots + c_{p,p-1}\mu_{p-1} + \mu_p &= 0 \\ c_{p0}\mu_1 + c_{p1}\mu_2 + \ldots + c_{p,p-1}\mu_p + \mu_{p+1} &= 0 \\ \cdot \qquad \cdot \qquad \cdots \qquad \cdot \qquad & \cdot \\ c_{p0}\mu_{p-1} + c_{p1}\mu_p + \ldots + c_{p,p-1}\mu_{2p-2} + \mu_{2p-1} &= 0. \end{aligned}$$

The solution may be expressed as a determinant in the usual way. Writing $\Delta^{(p-1)}$ in accordance with (22.21) and $\Delta^{(p)}_{pj}$ for the minor of the term in the last row and $(j+1)$th column in (22.21), we find

$$c_{pj} = \frac{\Delta^{(p)}_{pj}}{\Delta^{(p-1)}}. \qquad (22.32)$$

This expresses the c's in terms of the ascertainable constants μ. It follows that

$$P_p = \frac{1}{\Delta^{(p-1)}} \begin{vmatrix} \mu_0 & \mu_1 & \ldots & \mu_p \\ \mu_1 & \mu_2 & \ldots & \mu_{p+1} \\ \cdot & \cdot & \ldots & \cdot \\ \mu_{p-1} & \mu_p & \ldots & \mu_{2p-1} \\ 1 & X & \ldots & X^p \end{vmatrix} \qquad (22.33)$$

We notice in particular that, in virtue of the diagonal symmetry of $\Delta^{(p)}$, we have

$$c_{jk} = c_{kj}. \qquad (22.34)$$

22.13. In virtue of (22.31) we have

$$\Sigma (P_p^2) = \Sigma (x^p P_p)$$

and thus, from (22.33) on multiplying the last row and summing,

$$\Sigma (P_p^2) = \frac{n\Delta^{(p)}}{\Delta^{(p-1)}}. \qquad (22.35)$$

Similarly

$$\Sigma (y P_p) = \frac{n\Delta^{(p)}_p}{\Delta^{(p-1)}}. \qquad (22.36)$$

Finally, from (22.27)

$$b_p = \frac{\Delta_p^{(p)}}{\Delta^{(p)}}. \qquad (22.37)$$

Our problem is now solved. We have expressed all the unknowns in terms of calculable determinants.

We may note in passing that since the regression equation must remain covariant under a change of origin, all the coefficients b except b_0 are seminvariant, and the origin can thus be chosen at will. b_0 itself is the mean of the y-values.

22.14. Explicitly for the polynomials we have (taking $\mu_1 = 0$, $\mu_2 = 1$)—

$$P_0 = 1 \qquad (22.38)$$

$$P_1 = \frac{\begin{vmatrix} 1 & 0 \\ 1 & X \end{vmatrix}}{1} = X \qquad (22.39)$$

$$P_2 = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \mu_3 \\ 1 & X & X^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}} = X^2 - \mu_3 X - 1 \qquad (22.40)$$

$$P_3 = \frac{\begin{vmatrix} 1 & 0 & 1 & \mu_3 \\ 0 & 1 & \mu_3 & \mu_4 \\ 1 & \mu_3 & \mu_4 & \mu_5 \\ 1 & X & X^2 & X^3 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \mu_3 \\ 1 & \mu_3 & \mu_4 \end{vmatrix}}$$

$$= \frac{1}{\mu_4 - \mu_3^2 - 1}\{(\mu_4 - \mu_3^2 - 1)X^3 - (\mu_5 - \mu_4\mu_3 - \mu_3)X^2$$
$$+ (\mu_3\mu_5 - \mu_4^2 + \mu_4 - \mu_3^2)X + (\mu_5 - 2\mu_4\mu_3 + \mu_3)\} \qquad (22.41)$$

and so on. In particular, if the population is normal,

$$P_1 = X$$
$$P_2 = X^2 - 1$$
$$P_3 = X^3 - 3X, \text{ etc.,}$$

the polynomials in this case reducing to the Tchebycheff-Hermite functions (**6.20**) which we know to form an orthogonal system in the normal case.

Example 22.3. Ungrouped Data

Table 22.1 shows the relationship between the percentage loss in weight (Y) and the temperature (X) in a number of samples of soil. We require to find the regression of Y on X.

TABLE 22.1

Fitting of Curvilinear Regression for Ungrouped Data

(Data from J. R. H. Coutts, *J. Agr. Sci.*, **20**, 541.)

Percentage Loss in Weight. Y	Temperature (degrees F.). X
3·71	100
3·81	105
3·86	110
3·93	115
3·96	121
4·20	132
4·34	144
4·51	153
4·73	163
5·35	179
5·74	191
6·14	203
6·51	212
6·98	226
7·44	237
7·76	251

For the sums required we find—

$$n = 16,\ \Sigma(y) = 82{\cdot}97,\ \Sigma(y^2) = 459{\cdot}4363\ ;$$
$$\Sigma(x) = 2642,\ \Sigma(x^2) = 474{,}050,\ \Sigma(x^3) = 91{,}244{,}582\ ;$$
$$\Sigma(x^4) = 18{,}553{,}164{,}842,\ \Sigma(x^5) = 3{,}930{,}294{,}225{,}302\ ;$$
$$\Sigma(x^6) = 858{,}077{,}668{,}755{,}250\ ;\ \Sigma(yx) = 14{,}736{\cdot}19\ ;$$
$$\Sigma(yx^2) = 2{,}819{,}909{\cdot}45,\ \Sigma(yx^3) = 571{,}902{,}362{\cdot}11.$$

These can be run off fairly quickly on a machine. We have not bothered to take a different mean from those given, but in general a certain amount of arithmetic can be saved by so doing.

Considering first of all the straightforward approach of (22.24), we have for the straight line of closest fit,

$$\begin{vmatrix} Y & 1 & X \\ 82{\cdot}97 & 16 & 2642 \\ 14{,}736{\cdot}19 & 2642 & 474{,}050 \end{vmatrix} = 0,$$

reducing to

$$Y = 0{\cdot}660 + 2{\cdot}741\left(\frac{X}{100}\right). \qquad (22.42)$$

We have put $n\mu_j$ instead of μ_j in the second and third rows of the determinant, as we are clearly entitled to do.

Similarly we find for the second- and third-order parabolas—

$$Y = 3{\cdot}551 - 0{\cdot}929\left(\frac{X}{100}\right) + 1{\cdot}070\left(\frac{X}{100}\right)^2 \qquad (22.43)$$

$$Y = 7{\cdot}783 - 8{\cdot}940\left(\frac{X}{100}\right) - 5{\cdot}875\left(\frac{X}{100}\right)^2 - 0{\cdot}9189\left(\frac{X}{100}\right)^3 \qquad (22.44)$$

Fig. 22.1 shows the straight line and cubic fitted to the data by these means. An examination of the coefficients in the equations illustrates the point made above, that as successive terms are added to the polynomials the coefficients of all terms may alter very considerably.

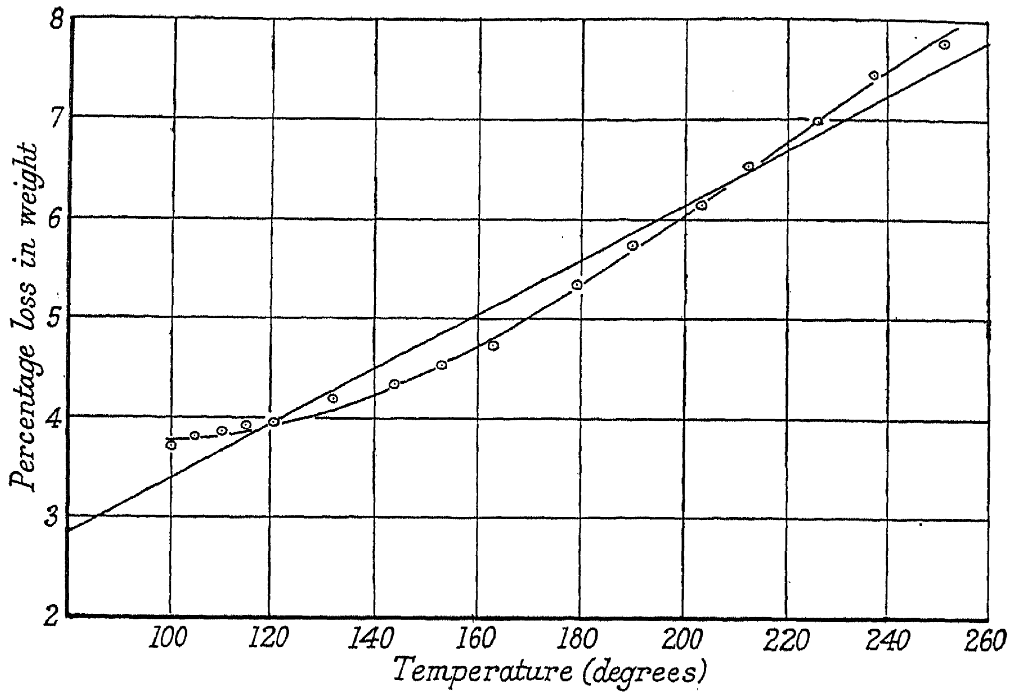


FIG. 22.1.—Straight Line and Cubic Parabola of Closest Fit to the Data of Table 22.1.

Consider now the alternative approach by the use of orthogonal polynomials. By the use of equations (22.33) we have

$$P_1 = \begin{vmatrix} 16 & 2642 \\ 1 & X \end{vmatrix} \Big/ 16$$
$$= X - 165{\cdot}125.$$

$$P_2 = \begin{vmatrix} 16 & 2642 & 474{,}050 \\ 2642 & 474{,}050 & 91{,}244{,}582 \\ 1 & X & X^2 \end{vmatrix} \Bigg/ \begin{vmatrix} 16 & 2642 \\ 2642 & 474{,}050 \end{vmatrix}$$
$$= \quad X^2 - 343{\cdot}137X + 27{,}032{\cdot}435.$$

$$P_3 = \frac{\begin{vmatrix} 16 & 2642 & 474{,}050 & 91{,}244{,}582 \\ 2642 & 474{,}050 & 91{,}244{,}582 & 18{,}553{,}164{,}842 \\ 474{,}050 & 91{,}244{,}582 & 18{,}553{,}164{,}842 & 3{,}930{,}294{,}225{,}302 \\ 1 & X & X^2 & X^3 \end{vmatrix}}{\begin{vmatrix} 16 & 2642 & 474{,}050 \\ 2642 & 474{,}050 & 91{,}244{,}582 \\ 474{,}050 & 91{,}244{,}582 & 18{,}553{,}164{,}842 \end{vmatrix}}$$
$$= \quad X^3 - 522{\cdot}940X^2 + 87{,}182{\cdot}434X - 4{,}605{,}047.$$

The b-coefficients are given by (22.37), the determinants in the numerator having been already tabulated in finding the P's. We have

$$b_0 = 5{\cdot}1856, \quad b_1 = \frac{2{\cdot}7409}{100}, \quad b_2 = \frac{1{\cdot}0695}{100^2}, \quad b_3 = -\frac{0{\cdot}91889}{100^3},$$

these being the values already found in arriving at (22.42) to (22.44). Thus

$$Y = 5{\cdot}1856 + \frac{2{\cdot}7409}{100}(X - 165{\cdot}125) + \frac{1{\cdot}0695}{100^2}(X^2 - 343{\cdot}137X + 27{,}032{\cdot}4)$$
$$- \frac{0{\cdot}91889}{100^3}(X^3 - 522{\cdot}940X^2 + 87{,}182{\cdot}4X - 4{,}605{,}047). \qquad (22.45)$$

If we stop at the second term we have

$$Y = 5{\cdot}1856 + \frac{2{\cdot}7409}{100}(X - 165{\cdot}125)$$
$$= 0{\cdot}660 + 2{\cdot}741\left(\frac{X}{100}\right),$$

which is the same as (22.42), as of course it must be. Similarly, if we stop at the third or fourth terms we find equations (22.43) or (22.44).

Now consider the fit of the regression line. We have from (22.35),

$$b_p^2\, \Sigma(P_p^2) = n\, b_p^2 \frac{\Delta^{(p)}}{\Delta^{(p-1)}} = b_p\, \Sigma(YP_p).$$

The determinants in this expression have already been evaluated in finding the regression line. Remembering that $\Sigma(y^2) = 459{\cdot}436$ we obtain the following :—

j.	b_j.	$n\, b_j^2 \frac{\Delta^{(j)}}{\Delta^{(j-1)}}$.	U (equation (22.28)).
0	5·1856	430·247	29·189
1	$2{\cdot}7409 \times 10^{-2}$	28·390	0·799
2	$1{\cdot}0695 \times 10^{-4}$	0·669	0·130
3	$-\ 0{\cdot}91889 \times 10^{-6}$	0·080	0·050

In calculations of this kind it is as well to take b_j to an extra place of decimals, as the value of U is rather sensitive to small errors of rounding up. Even so, the last figure in U is unreliable.

From the values of U it is clear that the fit is greatly improved by taking a quadratic term, and still further improved by adding the cubic term. How far a quartic term would improve matters cannot be decided without ascertaining the term. We have, however, not proceeded beyond the third degree because to do so would require moments of the eighth order. For a small population such as this, which in practical applications would be considered as a sample only, the errors in higher moments would probably be considerable.

The reader who works through the arithmetic of this example will find that there is about the same labour involved in either method. It is in the fitting of higher order terms that the method of orthogonal polynomials shows its superiority. In practical cases it is preferable to avoid the large numbers arising from the evaluation of determinants by a modification of the procedure given in **22.27** below.

Example 22.4. Grouped Data

In Example 14.1 (vol. I, p. 331) we considered the correlation between age and highest audible pitch in 3379 subjects and found the linear regressions. Let us take the work a stage further.

For the data of the table (X = age, Y = pitch) we find—

$$\Sigma(y) = -708; \quad \Sigma(y^2) = 8894; \quad \Sigma(yx) = -12{,}535;$$
$$\Sigma(x) = 2604; \quad \Sigma(x^2) = 47{,}392; \quad \Sigma(x^3) = 387{,}498;$$
$$\Sigma(x^4) = 4{,}842{,}172; \quad \Sigma(x^5) = 62{,}401{,}794; \quad \Sigma(x^6) = 883{,}576{,}012.$$

As a variation on the procedure of the previous example, we will convert these figures to moments about the mean (with Sheppard's corrections) and put them in standard measure. We find—

$$\mu_{01} = -0{\cdot}209{,}529; \quad \mu_{02} = 2{\cdot}504{,}904;$$
$$\mu_1 = 0{\cdot}770{,}642; \quad \mu_2 = 13{\cdot}348{,}229.$$

In standard measure the other moments are

$$\mu_3 = 1{\cdot}705{,}375; \quad \mu_4 = 6{\cdot}295{,}759;$$
$$\mu_5 = 20{\cdot}729{,}861; \quad \mu_6 = 78{\cdot}409{,}775.$$

We may now use equations (22.38), etc., direct, and find

$$P_0 = 1,\ P_1 = X,\ P_2 = X^2 - 1{\cdot}705X - 1,\ P_3 = X^3 - 3{\cdot}471X^2 - 0{\cdot}376X + 3{\cdot}560.$$

We now require the moments μ_{21} and μ_{31}. We find

$$\Sigma(yx^2) = -112{,}495$$
$$\Sigma(yx^3) = -1{,}399{,}639,$$

and hence, with Sheppard's corrections and in standard measure,

$$\mu_{21} = -1{\cdot}177{,}920 \qquad \mu_{31} = -4{\cdot}215{,}958.$$

We now find, from (22.37),

$$b_0 = 0$$
$$b_1 = -0{\cdot}613{,}626$$
$$b_2 = -0{\cdot}055{,}064$$
$$b_3 = 0{\cdot}010{,}205.$$

The regression line of the third degree is then

$$Y = -0{\cdot}6136X - 0{\cdot}0551\,(X^2 - 1{\cdot}705X - 1) + 0{\cdot}0102\,(X^3 - 3{\cdot}471X^2 - 0{\cdot}376X + 3{\cdot}560),$$

where the origin is at the mean and the units are in standard measure.

Standard Errors of Regression Coefficients

22.15. The standard errors of unknowns derived from least squares can be found by the use of a result due originally to Gauss. Suppose α_j is the true value of a_j and the residuals $y - \Sigma\alpha_j x^j$ are distributed normally with variance v. Writing $da_j = \alpha_j - a_j$, we have for the frequency function of the residuals—

$$f \propto \exp -\frac{1}{2v}\sum_s\Big(y - \sum_j \alpha_j x^j\Big)^2$$

$$\propto \exp -\frac{1}{2v}\Big\{\sum_s\Big(y - \sum_j a_j x^j\Big)^2 + \sum_s\sum_j\Big(da_j\, x^j\Big)^2\Big\}$$

($\sum_s$ denoting summation over the sample and $\sum_j$ over the values a_0 to a_p, and the cross-term vanishing because the a's are minimal values) ;

$$\propto \text{constant} \times \exp - \frac{1}{2v} \sum_s \sum_j (da_j\, x^j)^2$$

$$\propto \exp - \frac{1}{2v} \sum_s \sum_{j,k} (da_j\, da_k\, x^{j+k})$$

$$\propto \exp - \frac{n}{2v} \sum_{j,k} (da_j\, da_k\, \mu_{j+k}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.46)$$

In the limit, then, the deviations are distributed in the bivariate normal form, and from the results of **15.12** (vol. I, p. 376) it follows that

$$\text{var}\, a_j = \frac{\Delta_{jj}^{(p)}}{\Delta^{(p)}} \frac{v}{n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.47)$$

for the determinant whose terms are μ_{j+k} is in fact the determinant we have already defined as $\Delta^{(p)}$, and $\Delta_{jj}^{(p)}$ is the minor of the item in the jth row and column.

Now v is the variance of deviations from the theoretical regression line, and in terms of variations about the observed line we have, remembering the result of **18.17**—

$$\text{var}\, a_j = \frac{\Delta_{jj}^{(p)}}{\Delta^{(p)}} \cdot \frac{\text{var}\, e}{n - p - 1}. \quad . \quad . \quad . \quad . \quad . \quad (22.48)$$

Since the correlation ratio of y on x is given by

$$\text{var}\, e = \text{var}\, y\, (1 - \eta^2),$$

we have also

$$\text{var}\, a_j = \frac{\Delta_{jj}^{(p)}}{\Delta^{(p)}} \frac{(1 - \eta^2)\, \text{var}\, y}{n - p - 1}. \quad . \quad . \quad . \quad . \quad . \quad (22.49)$$

For large samples the replacement of n by $n - p - 1$ in the denominator is an unnecessary refinement.

22.16. For the case of orthogonal polynomials the results apply with a slight but important simplification. The coefficient b_j is the same as a_j if polynomials up to order j only are fitted, and hence, since $\Delta_{jj}^{(j)} = \Delta^{(j-1)}$ we have

$$\text{var}\, b_j = \frac{\Delta^{(j-1)}}{\Delta^{(j)}} \frac{(1 - \eta^2)\, \text{var}\, y}{n - j - 1}. \quad . \quad . \quad . \quad . \quad . \quad (22.50)$$

The same result follows by modifying (22.46), which for orthogonal polynomials becomes

$$f \propto \exp - \frac{1}{2v} \sum_j \left\{ \sum_s P_j^2\, (db_j)^2 \right\}, \quad . \quad . \quad . \quad . \quad . \quad (22.51)$$

showing that the b's are independently and normally distributed with variance

$$\text{var}\, b_j = \frac{v}{\sum P_j^2},$$

reducing to (22.50) in virtue of (22.35).

22.17. If the parent population is normal, $\eta = \rho$, and the determinants $\Delta^{(j)}$ can be evaluated explicitly in terms of the variance of x. In fact,

$$\frac{\Delta^{(j-1)}}{\Delta^{(j)}} = \frac{1}{j\,!\, (\text{var}\, x)^j}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.52)$$

and hence

$$\operatorname{var} b_j = \frac{1}{n - j - 1} \frac{(1 - \rho^2) \operatorname{var} y}{j\,!\,(\operatorname{var} x)^j}, \qquad (22.53)$$

or, in standard measure,

$$\operatorname{var} b_j = \frac{1}{n - j - 1} \cdot \frac{1 - \rho^2}{j\,!}. \qquad (22.54)$$

Equation (22.52) can be found by evaluating the determinants in the ordinary way, but it follows more simply from the consideration that $\frac{\Delta^{(j)}}{\Delta^{(j-1)}}$ is equal to $\frac{1}{n} \Sigma P_j^2$, which, in the normal case, is for large samples equal to $E(P_j^2) = j\,!\,(\operatorname{var} x)^j$ (**6.22.** vol. I, p. 147, with a change of scale).

22.18. The advantages of using orthogonal polynomials instead of powers of X are apparent in the forms taken by the standard errors of the coefficients a and b. The latter are independent of the order of the polynomial fitted and can be tested once and for all. The former do not possess this advantage. It seems preferable, therefore, as a matter of technique, to work with orthogonal polynomials throughout, whenever regressions of order higher than the first are likely to require investigation.

Example 22.5

Consider again the data of Example 22.4 (regression of highest audible pitch on age). We have there expressed the regression line in standard measure and in the orthogonal form, and may therefore use equation (22.50) in the form

$$\operatorname{var} b_1 = \frac{1 - \eta^2}{n} \frac{\Delta^{(0)}}{\Delta^{(1)}}$$

$$\operatorname{var} b_2 = \frac{1 - \eta^2}{n} \frac{\Delta^{(1)}}{\Delta^{(2)}}$$

$$\operatorname{var} b_3 = \frac{1 - \eta^2}{n} \frac{\Delta^{(2)}}{\Delta^{(3)}}.$$

(The sample number n is so large that we can ignore the element $-(j + 1)$ in the divisor.) The determinants required are already known, having been ascertained in the course of the work. We have

$$\frac{\Delta^{(0)}}{\Delta^{(1)}} = 1, \quad \frac{\Delta^{(1)}}{\Delta^{(2)}} = 0{\cdot}4189, \quad \frac{\Delta^{(2)}}{\Delta^{(3)}} = 0{\cdot}0985.$$

We also require η, which was found in Example 14.11 (vol. I, p. 352) to be $\eta_{yx} = 0{\cdot}6231$. Thus $1 - \eta^2 = 0{\cdot}6117$. We find

$$\operatorname{var} b_1 = \frac{1{\cdot}8104}{10^4}, \quad \operatorname{var} b_2 = \frac{0{\cdot}7584}{10^4}, \quad \operatorname{var} b_3 = \frac{0{\cdot}1783}{10^4}.$$

The values of the b's and their standard errors are then

Order.	b.	Standard Error.
1	− 0·6136	0·013
2	− 0·0551	0·0087
3	0·0102	0·0042

In all cases we should judge the coefficients significant, as being more than twice the standard error. Although, therefore, the second- and third-order terms are small and the regression is approximately linear, the deviation from linearity is not merely a chance effect.

Exact Significance Tests in the Normal Case

22.19. When the parent population is normal, more exact tests than those derived from the use of standard errors may be obtained. We have already seen (**14.21**, vol. I, p. 348) that a function dependent only on sample values and the first regression coefficient b_1 was distributed in " Student's " form. We proceed to generalise this result.

Consider in the first place the linear regression equation

$$Y = \bar{y} + b_1 (X - \bar{x}), \qquad (22.55)$$

and let β_1 be the population value of b_1 and σ_2^2 the variance of y in the population. Since the parent is normal, the variance of y for any fixed value of x is σ_2^2.

Our estimate of b_1 is

$$b_1 = \frac{\Sigma y (x - \bar{x})}{\Sigma (x - \bar{x})^2}, \qquad (22.56)$$

where summation takes place over the sample values. Thus *for fixed values of x* we have

$$\begin{aligned} \operatorname{var} b_1 &= \frac{\Sigma (x - \bar{x})^2 \operatorname{var} y}{\{\Sigma (x - \bar{x})^2\}^2} \\ &= \frac{\sigma_2^2}{\Sigma (x - \bar{x})^2}. \qquad (22.57) \end{aligned}$$

Thus, since the mean of the distribution of b_1 is β_1, we see that, for samples having the same x's as those observed, b_1 is normally distributed about mean β_1 with variance given by (22.57)—normally because it is a linear function of the y's which are themselves normal. Consequently,

$$\frac{(b_1 - \beta_1) \sqrt{\Sigma (x - \bar{x})^2}}{\sigma_2} \qquad (22.58)$$

is distributed normally about zero mean with unit variance.

If σ_2 were known this would provide a test of significance of b_1 in the ordinary way; but in fact σ_2 is not known and the substitution of an estimate distributed in the Type III form brings in the t-distribution in the usual way. We take as our estimator of σ_2 the function s, where

$$s^2 = \frac{1}{n-2} \Sigma (y - Y')^2, \qquad (22.59)$$

amd Y' represents the values " predicted " by the regression line, that is, the values

$$Y' = \bar{y} - b_1 (x - \bar{x}). \qquad (22.60)$$

Thus s^2 is based on the sum of squares of residuals. We shall show presently that s^2 is distributed in the Type III form with $n - 2$ degrees of freedom independently of $b_1 - \beta_1$. It follows that

$$t = \frac{(b_1 - \beta_1) \sqrt{\Sigma (x - \bar{x})^2} \sqrt{(n-2)}}{\sqrt{\Sigma (y - Y')^2}}$$

is distributed as " Student's " t with $\nu = n - 2$.

A given value β_1 may be tested accordingly. But we notice that the inference is a conditional one, that is to say, we are considering the distribution of t in a sub-population for which the x's are the same as those actually observed. (Cf. **21.47**.)

22.20. To establish the foregoing result we have to show that $\Sigma (y - Y')^2$, the sum of squares of residuals about the *observed* regression line, is distributed in the Type III form with $n - 2$ degrees of freedom. This is a particular case of a general theorem we shall prove at the beginning of the next chapter, but we will sketch an *ad hoc* proof here for the sake of completeness.

Since the population is normal, the deviations of y from the true regression line for fixed x's, $Y = \beta_0 + \beta_1 (X - \bar{x})$, where β_0 is the parent mean of y, is normal with variance σ_2^2. Now

$$(n - 2) \frac{s^2}{\sigma_2^2} = \frac{1}{\sigma_2^2} \Sigma (y - Y')^2 = \frac{1}{\sigma_2^2} \Sigma \{ y - b_0 - b_1 (x - \bar{x}) \}^2$$
$$= \frac{1}{\sigma_2^2} \Sigma \{ y - \beta_0 - \beta_1 (x - \bar{x}) - (b_0 - \beta_0) - (b_1 - \beta_1)(x - \bar{x}) \}^2.$$

The coefficients b_0 and b_1 were chosen so as to minimise this sum, and hence

$$(n - 2) \frac{s^2}{\sigma_2^2} = \frac{1}{\sigma_2^2} \Sigma \{ y - \beta_0 - \beta_1 (x - \bar{x}) \}^2 - \frac{n}{\sigma_2^2} (b_0 - \beta_0)^2 - \frac{(b_1 - \beta_1)^2}{\sigma_2^2} \Sigma (x - \bar{x})^2. \quad (22.61)$$

The first term is the sum of squares of n normal variates with zero mean and unit variance; the second is also such a variate, for it is the square of the deviation of the mean of y about its true value divided by the variance σ_2^2/n; and the third term is also such a variate, as shown above.

It does not follow immediately that $\dfrac{(n - 2) s^2}{\sigma_2^2}$ is distributed as the sum of squares of $n - 2$ normal variates in standard measure, for the constituent items might be correlated. Let us then find an orthogonal transformation to new variates $\xi_1 \ldots \xi_n$ linearly related to the n normal variates $y - \beta_0 - \beta_1 (x - \bar{x})$. These also will be normally and independently distributed. In particular (remembering that our summations refer to the y's and x's, but the latter are constant for our distributions), take

$$\xi_1 = \frac{1}{\sigma_2 \sqrt{n}} \Sigma \{ y - \beta_0 - \beta_1 (x - \bar{x}) \}$$
$$= \frac{\sqrt{n}}{\sigma_2} (b_0 - \beta_0)$$
$$\xi_2 = \frac{1}{\sigma_2} \Sigma \left[\frac{x - \bar{x}}{\sqrt{\Sigma (x - \bar{x})^2}} \{ y - \beta_0 - \beta_1 (x - \bar{x}) \} \right]$$
$$= \frac{1}{\sigma_2} (b_1 - \beta_1) \sqrt{\Sigma (x - \bar{x})^2}.$$

ξ_1 and ξ_2 are then normal variates in standard measure. Moreover they are orthogonal since

$$\Sigma \xi_1 \xi_2 = \frac{1}{\sigma_2^2 \sqrt{n}} \Sigma \frac{x - \bar{x}}{\sqrt{\Sigma (x - \bar{x})^2}}$$
$$= k \, \Sigma (x - \bar{x})$$
$$= 0.$$

Consequently our transformation exhibits the first term on the right in (22.61) as $\sum_{j=1}^{n} \xi_j^2$ and the second and third as ξ_1^2 and ξ_2^2. Thus the total is distributed as $\sum_{j=3}^{n} \xi_j^2$, which is the result required.

We may compare the result of **18.17**—in which we saw that the mean value of ε^2 was n, whereas that of e^2 was $n - p - 1$, one degree of freedom having been lost in the sum of squares of residuals for every constant estimated—and the approximate result of **21.20** in which χ^2 had to lose a degree for each constant fitted by maximum likelihood. Fundamentally all these results are different aspects of the same thing and rest on the fact that the variation of the sum of squares of normal variates in standard measure is spherically symmetric, so that a hyperplane in the sample space "cuts" the distribution in a spherically symmetric form of one lower degree of freedom.

Extension to Curvilinear Regression

22.21. The foregoing result can be extended without difficulty to the case when the regression is curvilinear. If

$$Y = b_0 P_0 + b_1 P_1 + \ldots + b_p P_p,$$

where the P's are orthogonal, then

$$b_j = \frac{\Sigma y_j P_j}{\Sigma P_j^2};$$

and we have also, for the variance of b_j when the x's are fixed,

$$\text{var } b_j = \frac{\sigma_2^2}{\Sigma P_j^2},$$

so that

$$\frac{(b_j - \beta_j)\sqrt{\Sigma P_j^2}}{\sigma_2}$$

is distributed normally with zero mean and unit variance. Taking as our estimate of σ_2^2

$$s^2 = \frac{1}{n-j-1}\Sigma(y - Y')^2,$$

we see, as before, that

$$t = \frac{(b_j - \beta_j)\sqrt{(n-j-1)}\sqrt{\Sigma P_j^2}}{\sqrt{\Sigma(y - Y')^2}} \quad . \quad . \quad . \quad . \quad (22.62)$$

is distributed as "Student's" t with $\nu = n - j - 1$ degrees of freedom.

It will be observed that in this and the previous section we have not assumed anything about the distribution in x-arrays. We have merely supposed that for any given x, y is normally distributed with constant variance.

Example 22.6

Consider again the soil data of Example 22.3. We found, for the cubic term in the parabola, a coefficient of $-0{\cdot}9189 \times 10^{-6}$. Is this significant ?

Here
$$b_j - \beta_j = -0{\cdot}9189 \times 10^{-6} \qquad \text{for } j = 3;$$
$$\sqrt{(n-j-1)} = \sqrt{(16-4)} = 3{\cdot}464.$$

We have already found $\Sigma(y - Y')^2 = U$, namely

$$U = 0{\cdot}050.$$

We further require ΣP_j^2 which has been obtained incidentally in the working of Example 22.3 and is equal to $9{\cdot}31525 \times 10^{10}$. Hence

$$-t = \frac{0{\cdot}9189 \times 10^{-6}\,(3{\cdot}464)\,3{\cdot}052 \times 10^5}{0{\cdot}2236}$$
$$\doteq 4{\cdot}3.$$

This is highly significant.

Case when the Independent Variate proceeds by Equal Steps

22.22. An important special case arises when the independent variate has values which are equidistant, as, for instance, in most time-series and in grouped data. If we take the interval between successive values of x as our unit, the variate-values may, by a suitable choice of origin, be taken as $0, 1, 2, \ldots n-1$. The various moment-functions μ_j entering into the expressions for polynomials, etc., may be written down once for all. Furthermore, this case lends itself to simpler summatory methods of forming the actual polynomial values and the residuals.

22.23. For a set of values $0, 1, 2, \ldots n-1$, we have

$$\Sigma(x) = \frac{n(n-1)}{2}, \qquad \Sigma(x^2) = \frac{n(n-1)(2n-1)}{6},$$

$$\Sigma(x^3) = \frac{n^2(n-1)^2}{4}, \text{ etc.}$$

Thus—

$$\mu_1 = \tfrac{1}{2}(n-1), \quad \mu_2 = \frac{n^2-1}{12}, \quad \mu_3 = 0, \text{ etc.}$$

From (22.38) and similar equations we then find

$$\left.\begin{aligned} P_1 &= X - \frac{n-1}{2} \\ P_2 &= \frac{X^2\mu_2 - X\mu_3 - \mu_2^2}{\mu_2} = P_1^2 - \frac{n^2-1}{12} \end{aligned}\right\} \quad . \quad . \quad . \quad (22.63)$$

and so on. The polynomials may be obtained more systematically as follows :—

We show first of all that

$$\sum_{j=0}^{p} \binom{n-1}{j} \frac{\Delta^j}{q+j} P_p = 0, \qquad q = 1, 2, \ldots p, \quad . \quad . \quad (22.64)$$

where Δ^j is the jth terminal difference of P_p and the x's range from 0 to $n-1$. In fact, from Newton's interpolation formula,

$$P_p = \sum_{j=0}^{p} \frac{X^{[j]}}{j!} \Delta^j P_p ; \quad . \quad . \quad . \quad . \quad . \quad (22.65)$$

and since the P's are orthogonal,

$$\sum_x (x+q-1)^{[q-1]} P_p = 0, \qquad q \leqslant p. \quad . \quad . \quad . \quad . \quad (22.66)$$

Substituting from (22.65), we find for the term in $\Delta^j P_p$—

$$\begin{aligned} \sum_x (x+q-1)^{[q+j-1]} \frac{\Delta^j}{j!} P_p &= \sum_x \{ (x+q)^{[q+j]} - (x+q-1)^{[q+j]} \} \frac{\Delta^j}{(j+q)j!} P_p \\ &= (n+q-1)^{[q+j]} \frac{\Delta^j}{(j+q)j!} P_p. \end{aligned}$$

Thus for all q from 1 to p we have

$$\begin{aligned} 0 &= \sum_{j=0}^{p} (n+q-1)^{[q+j]} \frac{\Delta^j}{(j+q)j!} P_p \\ &= \frac{(n+q-1)!}{(n-1)!} \Sigma \binom{n-1}{j} \frac{\Delta^j}{j+q} P_p, \qquad q \leqslant p \end{aligned}$$

whence follows (22.64). We now find functions obeying these conditions.

Consider

$$y = C(x+p)^{[p]}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.67)$$

This is a polynomial of degree p, and if for $x = 0, 1, \ldots p$ it assumes the values $y_0, \ldots y_p$ we have—

$$y(x) = C\, x^{[p+1]} \sum_{j=0}^{p} \frac{y_j\,(-1)^{p-j}}{j\,!\,(p-j)\,!\,(x-j)}, \qquad (22.68)$$

for this also is of degree p and has the right values at $x = 0, \ldots p$. Taking now

$$y_j = \frac{(n-1)\,!\,(p-j)\,!}{(n-j-1)\,!}\,(-1)^{p-j}\,\Delta^j P_p, \qquad (22.69)$$

we find that for $x = -q$

$$y(-q) = C\,(p+q)^{[p+1]}\,(-1)^p \sum_{j=0}^{p} \frac{(-1)^{p-j}}{j\,!\,(p-j)\,!}\,\frac{y_j}{q+j}$$

$$= C\,(-1)^p\,(p+q)^{[p+1]} \sum_{j=0}^{p} \binom{n-1}{j} \frac{\Delta^j}{q+j} P_p. \qquad (22.70)$$

Now from the definition of y this clearly vanishes for $-x = q = 1, \ldots p$, and thus (22.70) is zero. Comparing it with (22.64) we see that the conditions are satisfied if we give to y_j the value of Δ^j of (22.69), i.e.

$$\Delta^j P_p = \frac{(n-j-1)\,!}{(n-1)\,!\,(p-j)\,!}\,(-1)^{p-j}\,y_j$$

$$= C\,\frac{(n-j-1)\,!\,(p+j)\,!}{(n-1)\,!\,(p-j)\,!\,j\,!}\,(-1)^{p-j}. \qquad (22.71)$$

The constant C is evaluated by the fact that the coefficient of X^p in P_p is unity, giving $\Delta^p P_p = p\,!$ This gives

$$C = \frac{(p\,!)^2\,(n-1)\,!}{(2p)\,!\,(n-p-1)\,!}. \qquad (22.72)$$

Finally, substituting in (22.65), we find

$$P_p = \sum_{j=0}^{p} (-1)^{p-j}\,\frac{(p\,!)^2\,(p+j)\,!\,(n-j-1)\,!}{(2p)\,!\,(j\,!)^2\,(p-j)\,!\,(n-p-1)\,!}\,X(X-1)\ldots(X-j+1), \qquad (22.73)$$

where by convention the term $X^{[j]}$ is unity for $j = 0$. The first six polynomials are

$$\left.\begin{aligned}
P_1 &= X - \frac{n-1}{2} \\
P_2 &= P_1^2 - \frac{n^2-1}{12} \\
P_3 &= P_1^3 - \frac{3n^2-7}{20} P_1 \\
P_4 &= P_1^4 - \frac{3n^2-13}{14} P_1^2 + \frac{3\,(n^2-1)\,(n^2-9)}{560} \\
P_5 &= P_1^5 - \frac{5\,(n^2-7)}{18} P_1^3 + \frac{15n^4 - 230n^2 + 407}{1008} P_1 \\
P_6 &= P_1^6 - \frac{5\,(3n^2-31)}{44} P_1^4 + \frac{5n^4 - 110n^2 + 329}{176} P_1^2 \\
&\quad - \frac{5\,(n^2-1)\,(n^2-9)\,(n^2-25)}{14784}
\end{aligned}\right\} \qquad (22.74)$$

Four more values are given by Allan (1930), to whom the above derivation of (22.73) is due.

Values of the polynomials up to and including the fifth are given in Fisher and Yates' *Statistical Tables* up to $n = 52$.

22.24. We can now find an explicit expression for ΣP_p^2. Since the polynomials are orthogonal we have

$$\Sigma P_p^2 = \Sigma (x + p)^{[p]} P_p$$

which, by the argument resulting in (22.64), leads to

$$\Sigma P_p^2 = \sum_{j=0}^{p} \frac{(n+p)!}{j!\,(n-j-1)!} \frac{\Delta^j}{p+j+1} P_p.$$

Putting $q = p + 1$ in (22.67) and (22.70), we have

$$y(-q) = C(-1)^{[p]} = (-1)^p (2p+1)^{[p+1]} \sum_{j=0}^{p} \binom{n-1}{j} \frac{\Delta^j}{p+j+1} P_p,$$

whence, after a little rearrangement,

$$\Sigma \frac{(n+p)!}{j!\,(n-j-1)!} \frac{\Delta^j P_p}{p+j+1} = \frac{(p!)^2 (n+p)!}{(2p+1)!\,(n-1)!} C;$$

and thus, substituting for C from (22.72), we find

$$\Sigma P_p^2 = \frac{(p!)^4}{(2p)!\,(2p+1)!} n(n^2-1) \ldots (n^2-p^2). \qquad (22.75)$$

22.25. It is also possible to express the orthogonal polynomials in terms of central differences. We quote without proof the results (for details of which see Allan, 1930) :—

$$P_p = \frac{p!}{(p-\frac{1}{2})!} [\tfrac{1}{2}n]^p P_1 \Sigma \frac{(-1)^j (p-j-\frac{1}{2})!}{(p-2j)!\,j!\,2^{2j}} \frac{[P_1]^{p-2j-1}}{[\frac{1}{2}n]^{p-2j}} \qquad (22.76)$$

where

$$[x]^n = \frac{\{x + \frac{1}{2}(n-1)\}!}{\{x - \frac{1}{2}(n-1)\}!}. \qquad (22.77)$$

The series is summed from $j = 0$ until $2j > p$, when the denominator vanishes and $(p - \frac{1}{2})!$ is written for $\Gamma(p + \frac{1}{2})$ to preserve the factorial notation. In practice the polynomials for particular examples are not determined from (22.73) or (22.76) but by the use of tables, or by summation from differences in the manner of Example 22.9 below.

Example 22.7

For the fitting of a regression line in the case of equidistant intervals various methods are in use. A choice between them depends on the length of the series, the order of regression to which it is desired to go, and the computing resources at the investigator's disposal. We will illustrate two methods in this and the next example.

TABLE 22.2

Fitting of Regression Line by Orthogonal Polynomials—Equidistant x-intervals.

(1) Year.	(2) Variate. P_1	(3) Population (million). Y	(4) P_2	(5) $\frac{1}{6}P_3$	(6) $\frac{7}{12}P_4$
1811 . .	− 6	10·16	22	− 11	99
1821 . .	− 5	12·00	11	0	− 66
1831 . .	− 4	13·90	2	6	− 96
1841 . .	− 3	15·91	− 5	8	− 54
1851 . .	− 2	17·93	− 10	7	11
1861 . .	− 1	20·07	− 13	4	64
1871 . .	0	22·71	− 14	0	84
1881 . .	1	25·97	− 13	− 4	64
1891 . .	2	29·00	− 10	− 7	11
1901 . .	3	32·53	− 5	− 8	− 54
1911 . .	4	36·07	2	− 6	− 96
1921 . .	5	37·89	11	0	− 66
1931 . .	6	39·95	22	11	99

In Table 22.2, column 3 shows the population of England and Wales (in millions) for the years shown in column 1. These are at ten-yearly intervals, and the variate-values in units of 10 with origin at the mid-point of the range are given in column (2). These are the values of P_1.

The corresponding values of P_2, P_3 and P_4 are given in the last three columns. They may be calculated direct from (22.74), but are most conveniently taken direct from the Fisher–Yates tables.

We find, for $n = 13$,

$$\begin{aligned} \Sigma\, YP_1 &= 474{\cdot}77 \\ \Sigma\, YP_2 &= 123{\cdot}19 \\ \Sigma\, YP_3 &= -\ 39{\cdot}38 \times 6 = -\ 236{\cdot}28 \\ \Sigma\, YP_4 &= -\ 374{\cdot}30 \times \tfrac{12}{7} = -\ 641{\cdot}657{,}143, \end{aligned}$$

and, direct from the tables,

$$\Sigma\, P_1^2 = 182,\ \Sigma\, P_2^2 = 2002,\ \Sigma\, P_3^2 = 572 \times 36,$$
$$\Sigma\, P_4^2 = 68{,}068 \times (\tfrac{12}{7})^2.$$

Hence, from equations of the type $b_j = \dfrac{\Sigma\, YP_j}{\Sigma\, P_j^2}$, we find

$$b_1 = 2{\cdot}608{,}626, \quad b_2 = 0{\cdot}061{,}533{,}467, \quad b_3 = -\ 0{\cdot}011{,}474{,}359, \quad b_4 = -\ 0{\cdot}003{,}207{,}699$$

and the quartic curve is

$$Y - 24{\cdot}1608 = 2{\cdot}6086X + 0{\cdot}061{,}53\,(X^2 - 14) - 0{\cdot}011{,}47\,(X^3 - 25X)$$
$$-\ 0{\cdot}003{,}208\left(X^4 - \frac{247}{7}X^2 + 144\right) \qquad . \quad . \quad . \ (22.78)$$

We can now find the residuals for each term in this equation. We find

$$\begin{aligned} \Sigma\, Y^2 &= 8839{\cdot}9389 \\ \Sigma\, Y &= 314{\cdot}09. \end{aligned}$$

Hence the sum of squares of Y about the mean of Y,

$$\Sigma(Y - \bar{Y})^2 = 1251{\cdot}283.$$

Thus we have :—

		Residual Sum of Squares.
Original variation	1251·283	...
Contribution of first term $= b_1 \Sigma(YP_1)$	1238·497	12·786
Contribution of second term $= b_2 \Sigma(YP_2)$	7·580	5·206
Contribution of third term $= b_3 \Sigma(YP_3)$	2·711	2·495
Contribution of fourth term $= b_4 \Sigma(YP_4)$	2·058	0·437

For the *variance* of the residual elements we divide by the number of degrees of freedom $(n - j - 1)$ and obtain

Residual Sum of Squares.	Divisor.	Residual Variance.
12·786	11	1·162
5·206	10	0·521
2·495	9	0·277
0·437	8	0·055

Fig. 22.2 shows the data graphically with the cubic and quartic of closest fit.

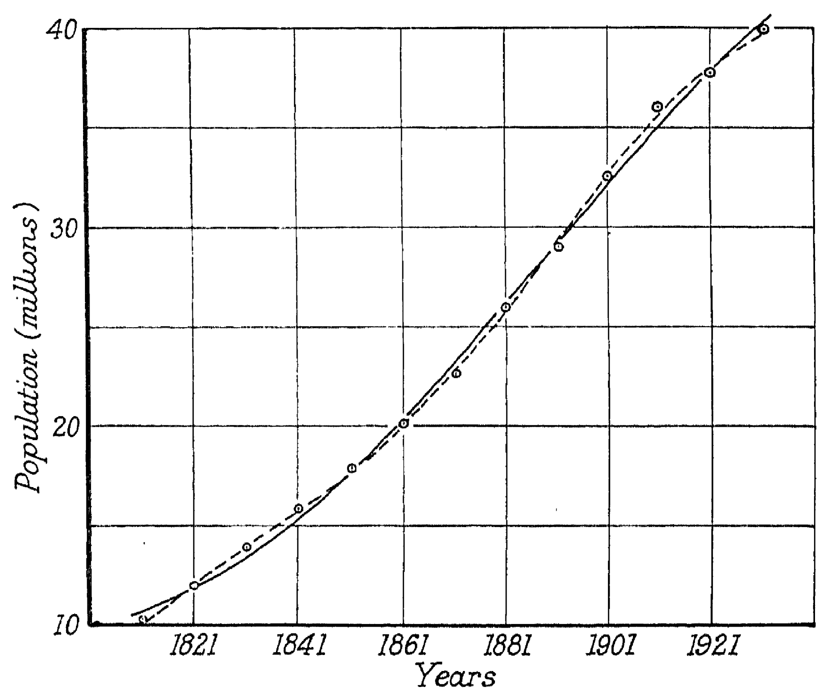


FIG. 22.2.—Cubic (full line) and Quartic (broken line) Parabolas fitted to the Data of Table 22.2.

The fit is evidently a good one, as is borne out by the smallness of the residual variance, but we must sound a warning as to the use of this polynomial. For interpolation in the variate range it would probably suit very well ; but for extrapolation outside the range it is dangerous unless there is good reason to suppose that the polynomial has some theoretical basis (which is not so). It would, for instance, be most unsafe to try and estimate the population in 1960 by inserting $X = 9$ in equation (22.78).

Example 22.8

In Chapter 3 it was seen that factorial moments can be derived by summatory processes. A somewhat similar method can be used to fit orthogonal polynomials. We will illustrate it on the data of the previous example.

TABLE 22.3

Fitting of Orthogonal Polynomials by Factorial Sums.

S_0	S_1	S_2	S_3
10·16	10·16	10·16	10·16
12·00	22·16	32·32	42·48
13·90	36·06	68·38	110·86
15·91	51·97	120·35	231·21
17·93	69·90	190·25	421·46
20·07	89·97	280·22	701·68
22·71	112·68	392·90	1094·58
25·97	138·65	531·55	1626·13
29·00	167·65	699·20	2325·33
32·53	200·18	899·38	3224·71
36·07	236·25	1135·63	4360·34
37·89	274·14	1409·77	5770·11
39·95	314·09	1723·86	7493·97
314·09	1723·86	7493·97	

In Table 22.3 the column headed S_0 gives the value of Y. The next column, headed S_1, gives the sums of the values in the first column proceeding from the top; and so for the columns headed S_2 and S_3.

Now construct the quantities

$$a_0 = \frac{1}{n} S_0 = \frac{314\cdot09}{13} = 24\cdot160{,}769$$

$$a_1 = \frac{2\,!}{n\,(n+1)} S_1 = \frac{2\,(1723\cdot86)}{182} = 18\cdot943{,}516$$

$$a_2 = \frac{3\,!}{n\,(n+1)\,(n+2)} S_2 = \frac{6\,(7493\cdot97)}{2730} = 16\cdot470{,}264$$

the general formula being

$$a_j = \frac{(j+1)\,!\,S_j}{n\,(n+1)\,.\,.\,.\,(n+j)}. \qquad (22.79)$$

Then obtain the quantities

$$\begin{aligned} a_0' &= a_0 = 24\cdot160{,}769 \\ a_1' &= a_0 - a_1 = 5\cdot217{,}253 \\ a_2' &= a_0 - 3a_1 + 2a_2 = 0\cdot270{,}749, \end{aligned}$$

the general formula being

$$a_p' = a_0 - \frac{p\,(p+1)}{(1\,!)^2\,2} a_1 + \frac{(p-1)\,(p)\,(p+1)\,(p+2)}{(2\,!)^2\,3} a_2 - \ldots \qquad (22.80)$$

Finally put
$$b_0 = a_0' = 24{\cdot}160{,}769$$
$$b_1 = \frac{6}{n-1} a_1' = \frac{6\,(5{\cdot}217{,}253)}{12} = 2{\cdot}608{,}626$$
$$b_2 = \frac{30}{(n-1)(n-2)} a_2' = \frac{30\,(0{\cdot}270{,}749)}{132} = 0{\cdot}061{,}534,$$
the general formula being
$$b_p = \frac{(2p+1)\,!}{(p\,!)^2} \frac{a_p'}{(n-1)\ldots(n-p)}. \qquad (22.81)$$
Then the b's are the coefficients of the orthogonal polynomials in the regression equation. The values we have found check with those of the previous example and the reader may care to work out b_3 and b_4 by the same method.

This process is due to R. A. Fisher and avoids the direct calculation of the values of the orthogonal polynomials. Its validity may be established by using equations (22.75) and (22.73), which give
$$b_p = \frac{\Sigma\, y P_p}{\Sigma P_p^2} = \frac{(2p\,!)\,(2p+1)\,!}{(p\,!)^4\, n\,(n^2-1)\ldots(n^2-p^2)} \Sigma\,(y P_p)$$
$$= \frac{(2p+1)\,!}{(p\,!)^2\,(n-1)\ldots(n-p)} \sum_j \frac{(-1)^{p-j}\,(p+j)\,!}{(j\,!)^2\,(p-j)\,!\,(j+1)} \frac{(n-j-1)\,!\,(j+1)\,\Sigma\, yx\ldots(x-j+1)}{(n-p-1)\,!\,n\ldots(n+p)}$$
The first part of the expression explains the coefficients in (22.81), the second part those in (22.80). The third part gives rise to (22.79) when it is remembered that the sums S are expressible as sums of factorials (cf. **3.10**, vol. I, p. 58), but the summation takes place from the top of the column.

Example 22.9

As a rule it is unnecessary to evaluate the polynomial at all the points for which data are given; but if the values are desired for comparison with observation they may be obtained by summatory processes from the differences.

The terminal differences themselves are obtainable simply from the quantities a_p' of the previous example. For a polynomial of the first degree we have
$$\left.\begin{aligned} \Delta Y &= -\frac{6}{n-1} a_1' \\ Y &= a_0' + 3a_1'. \end{aligned}\right\} \qquad (22.82)$$
For that of the second degree,
$$\left.\begin{aligned} \Delta^2 Y &= \frac{60}{(n-1)(n-2)} a_2' \\ \Delta Y &= -\frac{6}{n-1} (a_1' + 5a_2') \\ Y &= a_0' + 3a_1' + 5a_2'. \end{aligned}\right\} \qquad (22.83)$$
For the third degree,
$$\left.\begin{aligned} \Delta^3 Y &= \frac{-840}{(n-1)(n-2)(n-3)} a_3' \\ \Delta^2 Y &= \frac{60}{(n-1)(n-2)} (a_2' + 7a_3') \\ \Delta Y &= -\frac{6}{n-1} (a_1' + 5a_2' + 14a_3') \\ Y &= a_0' + 3a_1' + 5a_2' + 7a_3'. \end{aligned}\right\} \qquad (22.84)$$

The formulae for higher degrees are constructed on analogous lines, the multiplying factors for successive differences being given by

$$(-1)^p \frac{(p+1)(p+2)\ldots(2p+1)}{(n-1)(n-2)\ldots(n-p)}$$

and the coefficients of the a's by

Y	1	3	5	7	9	11	
ΔY		1	5	14	30	55	
$\Delta^2 Y$			1	7	27	77	etc.
$\Delta^3 Y$				1	9	44	
$\Delta^4 Y$					1	11	
$\Delta^5 Y$						1	

We leave the proof of these results to the reader.

For instance, for the data considered in the two previous examples we found, for the parabola of the second degree,

$$Y = 24{\cdot}160{,}8 + 2{\cdot}608{,}6X + 0{\cdot}061{,}533\,(X^2 - 14)$$
$$a_0' = 24{\cdot}160{,}769\,; \quad a_1' = 5{\cdot}217{,}253\,; \quad a_2' = 0{\cdot}270{,}749.$$

Hence, from (22.83),

$$\Delta^2 Y = \frac{60}{(n-1)(n-2)} a_2 = 0{\cdot}123{,}068$$

$$\Delta Y = -\frac{6}{n-1}(a_1' + 5a_2') = -3{\cdot}285{,}499$$

$$Y = a_0' + 3a_1' + 5a_2' = 41{\cdot}166{,}273.$$

We then build up the polynomial values as shown in Table 22.4. The second difference 0·123,068 is shown at the foot of column (2). Being a constant, it could have been written

TABLE 22.4

Calculation of Polynomial Values from Differences.

(1) Number of Term.	(2) Second Difference.	(3) First Difference.	(4) Polynomial Value.	(5) Observed Value.	(6) Difference (5)–(4)
1		− 1·808,68	9·863	10·16	0·297
2		− 1·931,75	11·795	12·00	0·205
3		− 2·054,82	13·849	13·90	0·051
4		− 2·177,88	16·027	15·91	− 0·117
5		− 2·300,95	18·328	17·93	− 0·398
6		− 2·424,02	20·752	20·07	− 0·682
7		− 2·547,09	23·299	22·71	− 0·589
8		− 2·670,16	25·969	25·97	0·001
9		− 2·793,23	28·763	29·00	0·237
10		− 2·916,29	31·679	32·53	0·851
11		− 3·039,36	34·718	36·07	1·352
12		− 3·162,43	37·881	37·89	0·009
13	0·123,068	− 3·285,499	41·166,27	39·95	− 1·216

all the way up, but to do so is a waste of time (and in practice, of course, we should not devote a separate column to it). The first difference is shown at the foot of column (3),

and the figures above it constructed by adding the second difference at each stage. The polynomial values themselves are compiled by adding the first differences to the value at the foot of the column, 41·166,27.

We have also shown the observed values and the difference between polynomial and observed values. The sum of squares of the latter is 5·204, agreeing within the margin of rounding-up error with the value for the sum of squares of residuals found in Example 22.7.

As an exercise the reader should work out the polynomial values for the third- and fourth-order polynomials and compare the sum of squares of residuals with the values of Example 22.7.

Multiple Curvilinear Regression

22.26. We considered the linear regression of one variate on a number of others in Chapters 14 and 15. There now remains the extension of our results to the curvilinear case.

The extension is very easy to carry out when we remember that in multiple linear regression there is no restriction on the degree of dependence among the "independent" variates. In particular, some of them may be functionally related, and more particularly still, one variate may be a power of another. It is thus clear that the process of fitting curved regression lines can be regarded as formally equivalent to that of fitting linear regressions. For instance, the fitting of

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + a_5 X_5$$

is equivalent to

$$Y = a_0 + a_1 X_1 + a_2 X_1^2 + a_3 Z_1 + a_4 Z_1^2 + a_5 Z_1^3,$$

the latter being a particular case of the former where X_2 is the square of X_1 (and their covariation accordingly complete) and similar relations exist between X_3, X_4 and X_5.

The case of curvilinear regression for a single variate, which has occupied the foregoing part of the chapter, could then have been treated by the methods of Chapter 15. We have discussed it afresh only because it is more easily dealt with by direct methods.

22.27. In multiple regression analysis it sometimes happens that, having worked out a regression equation, we wish either to take account of a new factor or to remove one which appears redundant. To avoid the necessity of solving a new set of determinantal equations the following device is useful:—

Consider the case of three independent variates measured from their mean

$$Y = b_1 X_1 + b_2 X_2 + b_3 X_3. \qquad (22.85)$$

In accordance with our general method the constants b are given by

$$\left.\begin{aligned} b_1 \Sigma(x_1^2) + b_2 \Sigma(x_1 x_2) + b_3 \Sigma(x_1 x_3) &= \Sigma(x_1 y) \\ b_1 \Sigma(x_1 x_2) + b_2 \Sigma(x_2^2) + b_3 \Sigma(x_2 x_3) &= \Sigma(x_2 y) \\ b_1 \Sigma(x_1 x_3) + b_2 \Sigma(x_2 x_3) + b_3 \Sigma(x_3^2) &= \Sigma(x_3 y) \end{aligned}\right\} \qquad (22.86)$$

Suppose now we replace the functions $\Sigma(xy)$ on the right by 1, 0, 0 and obtain the solutions $b_1 = c_{11}$, $b_2 = c_{12}$, $b_3 = c_{13}$; and similarly for replacement by 0, 1, 0 and 0, 0, 1, the solutions being written

$$\left.\begin{aligned} b_1 &= c_{11},\ c_{12},\ c_{13} \\ b_2 &= c_{12},\ c_{22},\ c_{23} \\ b_3 &= c_{13},\ c_{23},\ c_{33} \end{aligned}\right\}. \qquad (22.87)$$

Then the solution of (21.86) is

$$\left.\begin{aligned} b_1 &= c_{11}\,\Sigma(x_1 y) + c_{12}\,\Sigma(x_2 y) + c_{13}\,\Sigma(x_3 y)\\ b_2 &= c_{12}\,\Sigma(x_1 y) + c_{22}\,\Sigma(x_2 y) + c_{23}\,\Sigma(x_3 y)\\ b_3 &= c_{13}\,\Sigma(x_1 y) + c_{23}\,\Sigma(x_2 y) + c_{33}\,\Sigma(x_3 y) \end{aligned}\right\} \quad . \quad . \quad . \quad (22.88)$$

as is immediately evident on substitution. The values of the c's are those we have denoted earlier in the chapter by determinantal forms, e.g. $c_{jk} = \Delta^{(p)}_{jk}/\Delta^{(p)}$.

22.28. Now suppose that we wish to discard the variate x_3. From (22.86), with 1, 0, 0 written on the right, we find

$$c_{12} = -\frac{1}{\Delta}\begin{vmatrix} (11) & (13) & 1\\ (12) & (23) & 0\\ (13) & (33) & 0 \end{vmatrix} \quad . \quad . \quad . \quad . \quad . \quad (22.89)$$

where (jk) stands for $\Sigma(x_j x_k)$, and

$$\Delta = \begin{vmatrix} (11) & (12) & (13)\\ (12) & (22) & (23)\\ (13) & (23) & (33) \end{vmatrix} \quad . \quad . \quad . \quad . \quad . \quad (22.90)$$

There are similar expressions for the other c's. If the values of the constants when x_3 is removed are c'_{11}, c'_{12}, c'_{22} we shall have

$$c'_{11} = -\frac{1}{\Delta'}\begin{vmatrix} (12) & 1\\ (22) & 0 \end{vmatrix}, \qquad c'_{12} = \frac{1}{\Delta'}\begin{vmatrix} (11) & 1\\ (12) & 0 \end{vmatrix} \text{ etc.} \quad . \quad (22.91)$$

where

$$\Delta' = \begin{vmatrix} (11) & (12)\\ (12) & (22) \end{vmatrix} \quad . \quad . \quad . \quad . \quad . \quad (22.92)$$

Now we have

$$\frac{c_{13}\,c_{23}}{c_{33}} = \frac{\begin{vmatrix} (11) & (12) & 1\\ (12) & (22) & 0\\ (13) & (23) & 0 \end{vmatrix}\begin{vmatrix} (11) & (12) & 0\\ (12) & (22) & 1\\ (13) & (23) & 0 \end{vmatrix}}{\Delta\begin{vmatrix} (11) & (12) & 0\\ (12) & (22) & 0\\ (13) & (23) & 1 \end{vmatrix}}$$

$$= -\frac{\begin{vmatrix} (12) & (22)\\ (13) & (23) \end{vmatrix}\begin{vmatrix} (11) & (12)\\ (13) & (23) \end{vmatrix}}{\Delta\Delta'}$$

Thus

$$c_{12} - \frac{c_{13}\,c_{23}}{c_{33}} = \frac{c_{12}\,c_{33} - c_{13}\,c_{23}}{c_{33}}$$

$$= -\frac{\begin{vmatrix} (12) & (23)\\ (13) & (33) \end{vmatrix}\begin{vmatrix} (11) & (12)\\ (12) & (22) \end{vmatrix} - \begin{vmatrix} (12) & (22)\\ (13) & (23) \end{vmatrix}\begin{vmatrix} (11) & (12)\\ (13) & (23) \end{vmatrix}}{\Delta\Delta'}$$

$$= -\frac{(12)\,\Delta}{\Delta\Delta'}$$

$$= c'_{12}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.93)$$

Similarly

$$c'_{11} = c_{11} - \frac{c_{13}^2}{c_{33}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.94)$$

$$c'_{22} = c_{22} - \frac{c_{23}^2}{c_{33}}. \quad . \quad . \quad . \quad . \quad . \quad (22.95)$$

This gives us the new c's in terms of the old. Denoting similarly the new b's by primes, we have

$$\begin{aligned} b_1 - b'_1 &= (c_{11} - c'_{11}) \Sigma (x_1 y) + (c_{12} - c'_{12}) \Sigma (x_2 y) + c_{13} \Sigma (x_3 y) \\ &= \frac{1}{c_{33}} \{ c_{13}^2 \Sigma (x_1 y) + c_{13} c_{23} \Sigma (x_2 y) + c_{13} c_{33} \Sigma (x_3 y) \} \\ &= \frac{c_{13} b_3}{c_{33}}. \end{aligned}$$

Hence we have

$$\left. \begin{aligned} b'_1 &= b_1 - \frac{c_{13} b_3}{c_{33}} \\ b'_2 &= b_2 - \frac{c_{23} b_3}{c_{33}} \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (22.96)$$

expressing the new constants in terms of the old and the known constants c.

Finally, the contribution to the sum of squares due to the variate x_3 is

$$\begin{aligned} & b_1 \Sigma (x_1 y) + b_2 \Sigma (x_2 y) + b_3 \Sigma (x_3 y) - b'_1 \Sigma (x_1 y) - b'_2 \Sigma (x_2 y) \\ &= \frac{c_{13}}{c_{33}} b_3 \Sigma (x_1 y) + \frac{c_{23}}{c_{33}} b_3 \Sigma (x_2 y) + b_3 \Sigma (x_3 y) \\ &= \frac{b_3^2}{c_{33}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.97) \end{aligned}$$

22.29. Generally, if there are p independent variates the equations for the b's are

$$\begin{aligned} b_1 \Sigma (x_1^2) \quad &+ b_2 \Sigma (x_1 x_2) + \ . \ . \ . + b_p \Sigma (x_1 x_p) = \Sigma (y x_1) \\ &\cdot \qquad \cdot \qquad \cdot \qquad \cdot \\ b_1 \Sigma (x_1 x_p) &+ b_2 \Sigma (x_2 x_p) + \ . \ . \ . + b_p \Sigma (x_p^2) \quad = \Sigma (y x_p). \end{aligned}$$

If x_p is omitted the equations become $(p - 1)$ in number in variables $b'_1 \ . \ . \ . \ b'_{p-1}$. Subtracting from these the first $(p - 1)$ of the above equations we find $(p - 1)$ equations, typified by

$$(b'_1 - b_1) \Sigma (x_1 x_j) + (b'_2 - b_2) \Sigma (x_2 x_j) + \ . \ . \ . + (b'_{p-1} - b_{p-1}) \Sigma (x_{p-1} x_j) - b_p \Sigma (x_j x_p) = 0 \quad (22.98)$$

But these equations are the same as those for the coefficients $c_{1p} \ . \ . \ . \ c_{pp}$ with $(b'_1 - b_1)$ in place of c_{1p}, etc., and $- b_p$ in place of c_{pp}. Hence

$$\frac{b'_1 - b_1}{- b_p} = \frac{c_{1p}}{c_{pp}},$$

or

$$b'_1 - b_1 = - \frac{c_{1p} b_p}{c_{pp}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (22.99)$$

Similarly it will be found that

$$\left.\begin{aligned} c'_{11} - c_{11} &= -\frac{c_{1p}^2}{c_{pp}} \\ c'_{12} - c_{12} &= -\frac{c_{1p}\,c_{2p}}{c_{pp}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (22.100)$$

with similar equations for the other c's.

22.30. Somewhat similar results apply when a variate is added. If primes again refer to new coefficients when x_q is added, we have, as above—

$$\left.\begin{aligned} b'_1 - b_1 &= \frac{c'_{1q}\,b'_q}{c'_{qq}} \\ c'_{11} - c_{11} &= \frac{c'^2_{1q}}{c'_{qq}} \\ c'_{12} - c_{12} &= \frac{c'_{1q}\,c'_{2q}}{c'_{qq}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (22.101)$$

In order to use these equations to adjust the constants we require $c'_{1q} \ldots c'_{qq}$ and b'_q. By writing down the equations satisfied by $c_{11} \ldots c_{1p}$ and subtracting the corresponding equations in $c'_{11} \ldots c'_{1q}$, we get p equations such as

$$(c'_{11} - c_{11})\,\Sigma\,(x_1\,x_j) + \ldots + (c'_{1p} - c_{1p})\,\Sigma\,(x_j\,x_p) = -\,c'_{1q}\,\Sigma\,(x_j\,x_q).$$

These are the same as the equations in $b_1 \ldots b_q$ with $-\,c'_{1q}\,\Sigma\,(x_j\,x_q)$ instead of $\Sigma\,(x_q\,y)$ on the right, and hence

$$c'_{1p} - c_{1p} = -\,c'_{1q}\sum_{j=1}^{p} c_{pj}\,\Sigma\,(x_j\,x_q).$$

Thus, using (22.101),

$$\frac{c'_{pq}}{c'_{qq}} = -\sum_{j=1}^{p} c_{pj}\,\Sigma\,x_j\,x_q. \quad . \quad . \quad . \quad . \quad (22.102)$$

The last of the equations satisfied by c'_{qq} is

$$c'_{1q}\,\Sigma\,(x_q\,x_1) + \ldots + c'_{pq}\,\Sigma\,(x_q\,x_p) + c'_{qq}\,\Sigma\,(x_q^2) = 1.$$

Substituting for c'_{1q}, etc., in terms of c'_{qq}, we get

$$c'_{qq}\left\{\Sigma\,(x_q^2) - \sum_{j,k=1}^{p} c_{jk}\,\Sigma\,(x_j\,x_q)\,\Sigma\,(x_k\,x_q)\right\} = 1. \quad . \quad . \quad . \quad (22.103)$$

This gives c'_{qq}, and $c'_{1q} \ldots c'_{pq}$ are derivable from (22.102). The other constants then result from (22.101).

Cochran (1938a), to whom this proof is due, says that the elimination of two variates is best carried out in two stages of one each; that where one variate is eliminated the method is quicker than re-solving the regression equations, except where there are only two independent variates in the first instance; and that if two variates are being eliminated the method is quicker if the original number of independent variates is six or more. For the addition of variates the method is in all cases more expeditious than re-solving the regression equations.

Example 22.10 (Cochran, 1938*a*)

In a study of the effect of weather factors on the number of noctuid moths per night caught in a light-trap, regressions were worked out on X_1 (minimum night temperature), X_2 (the maximum temperature of the previous day), X_3 (the average speed of the wind during the night), and X_4 (the amount of rain during the night). The dependent variate was $\log(1 + n)$, where n was the number of moths.

It was subsequently decided to investigate the effect of cloudiness, measured on a conventional scale as the percentage of starlight obscured by clouds in a night sky camera. This is the new variate X_5.

The quantities c_{jk} for the first four variates were :—

	X_1	X_2	X_3	X_4
X_1	+ 0·105,423,56	− 0·041,946,20	− 0·096,067,09	− 0·018,490,96
X_2	...	+ 0·086,038,69	+ 0·033,172,71	+ 0·012,903,58
X_3	...	...	+ 0·572,652,01	+ 0·008,116,62
X_4	...	...	...	+ 0·062,275,32

and the sums $\Sigma(x_j x_5)$ were

$$\Sigma(x_1 x_5) = -4{\cdot}867, \quad \Sigma(x_2 x_5) = +0{\cdot}206, \quad \Sigma(x_3 x_5) = -0{\cdot}5446,$$
$$\Sigma(x_4 x_5) = -5{\cdot}42, \quad \Sigma(x_5^2) = 7{\cdot}87.$$

We then find from (22.103)

$$c'_{55} = +0{\cdot}210{,}133{,}14,$$

and from (22.102)

$$\frac{c'_{15}}{c'_{55}} = +0{\cdot}369{,}198{,}24 \quad \frac{c'_{25}}{c'_{55}} = -0{\cdot}133{,}872{,}86 \quad \frac{c'_{35}}{c'_{55}} = -0{\cdot}118{,}533{,}74$$
$$\frac{c'_{45}}{c'_{55}} = +0{\cdot}249{,}298{,}91,$$

so that the new c's are given by (22.101) as

	X_1	X_2	X_3	X_4	X_5
X_1	0·134,066,25	− 0·052,332,16	− 0·105,263,03	+ 0·000,849,84	+ 0·077,580,79
X_2	...	+ 0·089,804,68	+ 0·036,507,20	+ 0·005,890,52	− 0·028,131,12
X_3	...	...	+ 0·575,604,43	+ 0·001,907,12	− 0·024,907,87
X_4	...	...	...	+ 0·075,335,08	+ 0·052,385,96
X_5	...	...	...	...	+ 0·210,133,14

The original regression coefficients were

$$b_1 = +0{\cdot}198{,}140{,}7 \quad b_2 = +0{\cdot}038{,}528{,}4 \quad b_3 = -0{\cdot}508{,}649{,}2,$$
$$b_4 = +0{\cdot}031{,}848{,}2.$$

We now find

$$b'_5 = \sum_{j=1}^{5}\{c'_{j5}\,\Sigma(x_j y)\}$$
$$= -0{\cdot}227{,}149{,}6,$$

and from (22.101) we then have

$$b'_1 = +0{\cdot}114{,}277{,}5 \quad b'_2 = +0{\cdot}068{,}937{,}6 \quad b'_3 = -0{\cdot}481{,}724{,}3,$$
$$b'_4 = -0{\cdot}024{,}779{,}9.$$

As usual we have retained more figures than are necessary, in order to avoid cumulating errors and to facilitate the detection of computational slips.

22.31. The constants c found in the foregoing method have a further use: they give the standard errors of the regression coefficients and provide some of the functions required in more exact tests based on the t-distribution. If, measuring y about the mean, we have

$$Y = b_1 X_1 + b_2 X_2 + \ldots + b_p X_p,$$

then there are p equations of the kind:

$$\Sigma (x_1 y) = b_1 \Sigma x_1^2 + b_2 \Sigma (x_1 x_2) + \ldots + b_p \Sigma (x_1 x_p),$$

and thus, recalling the definition of the c's, we have

$$b_1 = c_{11} \Sigma (x_1 y) + c_{12} \Sigma (x_2 y) + \ldots + c_{1p} \Sigma (x_p y).$$

Thus, *for fixed values of the x's*,

$$\begin{aligned} \text{var } b_1 &= \text{var } y \left(\sum_{j,\, k} c_{1j} c_{1k} x_j x_k \right) \\ &= c_{11} \text{ var } y, \qquad (22.104) \end{aligned}$$

and so for the other b's.

For large samples var y may be taken to be the estimated variance

$$\frac{1}{n - p - 1} \Sigma (y - \bar{y})^2.$$

If the sample is small and it is desired to make a more accurate test, then we have, by an extension of **22.21**, that

$$t = \frac{(b_j - \beta_j) \sqrt{(n - p - 1)}}{\sqrt{\Sigma (y - \bar{y})^2} \sqrt{c_{jj}}} \qquad (22.105)$$

is distributed in "Student's" form with $\nu = n - p - 1$ degrees of freedom.

22.32. As a final comment we may emphasise that regression equations are only polynomials fitted to the means of arrays, and consequently that if the scatter about those means is substantial they are not very reliable as estimators (though they may be better than other methods). The comment would hardly be necessary were it not for a tendency to use the equations somewhat uncritically for purposes of *prediction*. The point assumes even greater importance when attempts are made to estimate the dependent variate for values of the independent variates outside the range on which the regressions are based; or again, if the observations are distributed over time so that the population may be changing while the sample is being drawn. The technique of regression analysis is undoubtedly useful in many fields, but—as with many other statistical techniques—the careful investigator will apply it with a certain amount of self-discipline.

NOTES AND REFERENCES

The theory of curvilinear regression was studied by Karl Pearson (1905). Orthogonal polynomials had been considered, and the essential problems solved, by Tchebycheff as far back as 1857, but their use in statistics was not fully appreciated until about sixty years later. Pearson gave in 1921 the general formulae for fitting curved regression lines up to the fourth order. Neyman (1926) pointed out the elegance of the determinantal approach.

From about 1920 onwards there may be discerned two main lines of development. The Scandinavian school, led by Wicksell, has developed the analytical theory of regression —see Wicksell (1917*b*, 1933, 1934*b*) and a useful memoir by W. Andersson (1932). The

second line, followed by Fisher, Aitken and others, has been concerned with the fitting of regression curves to arithmetical data and exact significance tests—see Fisher's papers of 1921*b*, 1922*b*, 1924*b*, 1926*a*, a paper by Allan (1930), and three papers by Aitken (1933*a*, *b*, *c*). The literature on orthogonal polynomials is now very large.

For some illustrative material, see K. Pearson (1905), Andersson (1932), and Pretorius (1930). See also references to Chapters 14 and 15.

EXERCISES

22.1. Show that the regression of y on the variance of x (the scedastic curve) is given by

$$Y = \sum_0^\infty (-1)^j \frac{\kappa_{j2} + \lambda_j}{j!} \frac{D^j g(X)}{g(X)} - \left[\sum_0^\infty \frac{(-1)^j}{j!} \sum_{s=0}^{j} \binom{j}{s} \kappa_{s1}\, \kappa_{j-s,1} \frac{D^s g(X)}{g(X)} \frac{D^{j-s} g(X)}{g(X)}\right]^2$$

where
$$\Sigma\left(\frac{\lambda_j t^j}{j!}\right) = \left(\Sigma \frac{\kappa_{j1} t^j}{j!}\right)^2$$

(Wicksell, 1934*b*.)

22.2. Show that if the regression of y on the mean of x is linear, then from (22.11)

$$\phi(t_1) \sum_0^\infty \frac{\kappa_{j1} t^j}{j!}$$

is a linear function of $\phi(t_1)$ and $\dfrac{d}{dt_1}\phi(t_1)$. Hence that

$$\kappa_{j1}\,\kappa_{20} = \kappa_{11}\,\kappa_{j+1,0}$$

(Wicksell, 1934*b*.)

22.3. Show that if the marginal distribution of a bivariate distribution is of the Gram-Charlier Type A :

$$f = \alpha(x)\{1 + a_3 H_3 + a_4 H_4 + \ldots\}$$

the regression of y on x is

$$Y = \frac{\displaystyle\sum_{j=0}^\infty \sum_{k=0}^\infty \frac{\kappa_{j1}}{j!} a_k H_{j+k}(X)}{1 + \displaystyle\sum_{j=3}^\infty a_j H_j(X)}$$

(Wicksell, 1917*b*.)

22.4. Transforming the orthogonal polynomials of (22.74) to a new variate $\xi = X - \dfrac{n-1}{2}$, note that $P_p - \xi P_{p-1}$ is a numerical multiple of P_{p-2}, say λP_{p-2}. Show that

$$\lambda = -\frac{\Sigma P_{p-1}^2}{\Sigma P_{p-2}^2},$$

and deduce the recurrence relation,

$$P_p = \xi P_{p-1} - \frac{(p-1)^2\{n^2 - (p-1)^2\}}{4(2p-1)(2p-3)} P_{p-2}.$$

(Allan, 1930. The relation is due to Tchebycheff.)

22.5. A regression line

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 X^3 + a_4 X^4$$

is fitted to normal data and the number of observations N is large. If r is the correlation between the variates and $c = \frac{\mu_1'^2}{\mu_2}$ (the moments referring to the x-variate), show that

$$\text{var } a_0 = \frac{\text{var } y}{24N}(45 + 30c^2 - 8c^3 + c^4)(1 - r^2)$$

$$\text{var } a_1 = \frac{\text{var } y}{6N\mu_2}(15 + 30c - 15c^2 + 4c^3)(1 - r^2)$$

$$\text{var } a_2 \doteq \frac{\text{var } y}{2N\mu_2^2}(4 - 3c + 3c^2)(1 - r^2)$$

$$\text{var } a_3 = \frac{\text{var } y}{6N\mu_2^3}(1 + 4c)(1 - r^2)$$

$$\text{var } a_4 = \frac{\text{var } y}{24N\mu_2^4}(1 - r^2).$$

(Andersson, 1932.)

22.6. In the notation of **22.31** show that

$$\text{cov }(b_1 b_2) = c_{12} \text{ var } y$$

and hence show how to test the difference of two coefficients in a regression equation.

22.7. Show how to derive a test of the significance of the difference of corresponding regression coefficients in two equations derived from independent samples, based on the result of **21.26**.

CHAPTER 23

THE ANALYSIS OF VARIANCE—(1)

23.1. At various points in this book we have encountered in different guises the result that the sum of squares of a set of observations about their mean can be represented as the sum of two independent sums of squares, each of which provides an estimate of the parent variance; and that their ratio provides a test of homogeneity, at least when the parent is normal. We now proceed to study in more detail a method of statistical analysis with considerable generality which springs from this result. In view of the complexity of the general case we shall begin by considering simpler cases under somewhat restrictive conditions and shall extend our results stage by stage.

One-way Classification

23.2. Suppose we have a set of variate-values divided into p families:

$$\begin{array}{lll} x_{11} & x_{21} \;.\;.\;.\;.\;.\; & x_{n_1 1} \\ x_{12} & x_{22} \;.\;.\;.\;.\;.\; & x_{n_2 2} \\ \cdot & \cdot \;\; \cdot & \cdot \;\; \cdot \\ x_{1p} & x_{2p} \;.\;.\;.\;.\;.\; & x_{n_p p}. \end{array}$$

Denoting by $\bar{x}$ the mean of the whole set and by $\bar{x}_j$ the mean of the values in the jth family, we have the identity

$$\begin{aligned} \sum_{i,j} (x_{ij} - \bar{x})^2 &= \sum_{i,j} (x_{ij} - \bar{x}_i + \bar{x}_j - \bar{x})^2 \\ &= \sum_{i,j} (x_{ij} - \bar{x}_j)^2 + \sum_{i,j} (\bar{x}_j - \bar{x})^2, \end{aligned} \qquad (23.1)$$

since the cross-product term $2\sum_{i,j} (x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x})$ vanishes. We may also write this as

$$\sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} (x_{ij} - \bar{x}_j)^2 + \sum_{j} n_j (\bar{x}_j - \bar{x})^2, \qquad (23.2)$$

where n_j is the number of members in the jth family.

It will also be convenient, from the point of view of a later generalisation, to write the mean of the jth family as $x_{.j}$ and that of the whole as $x_{..}$, the periods in the subscripts showing which factor is being averaged. We have then the alternative form

$$\sum_{i,j} (x_{ij} - x_{..})^2 = \sum_{i,j} (x_{ij} - x_{.j})^2 + \sum_{j} n_j (x_{.j} - x_{..})^2 \qquad (23.3)$$

23.3. The problem we shall discuss in connection with families of values of this type takes some such form as the following: the members of each family are randomly chosen from some parent population corresponding to that family. The populations themselves are, as a rule, defined by some prior system of classification given among the data of the problem, e.g. they might be different varieties of wheat, the x's being the yields of the varieties grown under similar conditions, or they might be defined by income levels and the x's the expenditure on food of a sample chosen from the different income groups. We now ask: is there any evidence that the factor measured by x varies significantly from

family to family ? Alternatively, can the data be regarded as homogeneous, i.e. as emanating from populations which are identical so far as concerns the factor measured by x ? Further, when the question of significance is decided, how can we estimate the variation of x in families or groups of families, and how can we estimate the magnitude of any differences which exist ?

23.4. We will assume, until further notice, that within each family the variation is normal with variance v, and that v is the same for each family. In later sections we shall endeavour to remove these rather restrictive conditions. On our present hypothesis the populations corresponding to the different families can differ, if at all, only in their means, and our first question is whether the sample values afford any evidence of such differences.

Let us take as our hypothesis that the parent populations have a common mean m. Then we recall the following facts :—

(1) The sum $\frac{1}{v}\Sigma(x_{ij} - x_{..})^2$ is distributed in the Type III form of χ^2 with $N - 1 = \Sigma_j (n_j) - 1$ degrees of freedom, that is to say as the sum of squares of $N - 1$ independent normal variates with zero mean and unit variance.

(2) In any given family $x_{.j}\sqrt{\frac{n_j}{v}}$ is distributed normally with unit variance about mean m, and is independent of the sum $\frac{1}{v}\Sigma_i (x_{ij} - x_{.j})^2$ which is itself distributed as χ^2 with $n_j - 1$ degrees of freedom.

Since on our hypothesis the observations may be regarded as a single sample from the same population, it follows that

$$\left.\begin{array}{lll}\frac{1}{v}\sum_{i,j}(x_{ij} - x_{..})^2 & \text{is distributed as } \chi^2 \text{ with} & N - 1 \text{ d.f.} \\ \frac{1}{v}\sum_{i,j}(x_{ij} - x_{.j})^2 & \quad\text{,,} \qquad\qquad\text{,,} & \Sigma(n_j - 1) = N - p \text{ d.f.} \\ \frac{1}{v}\Sigma_j n_j (x_{.j} - x_{..})^2 & \quad\text{,,} \qquad\qquad\text{,,} & p - 1 \text{ d.f.}\end{array}\right\} \qquad (23.4)$$

The only statement requiring any proof is the last. It may be proved directly (see Exercise 23.1), but we shall deduce it as the corollary of a general theorem due to R. A. Fisher which will often be required in this chapter.

23.5. Suppose we have q variates $x_1 \ldots x_q$ which are independently and normally distributed with unit variance about the same mean, which we may assume to be zero. Put

$$\zeta_r = \sum_{s=1}^{q} \lambda_{rs} x_s, \qquad r = 1 \ldots q. \quad . \quad . \quad . \quad . \quad (23.5)$$

If we choose the coefficients λ so that

$$\left.\begin{array}{ll}\Sigma_s \lambda_{rs}\lambda_{ts} = 1 & r = t \\ \qquad\quad = 0 & r \neq t\end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (23.6)$$

then each ζ is distributed normally with unit variance independently of the others. There

are q^2 coefficients λ, and the equations (23.6) impose $\frac{1}{2}q(q+1)$ conditions on them, so that the λ's can always be found in a multiplicity of ways. In effect they correspond to the rotation of orthogonal co-ordinate axes in a q-dimensional space.

Now suppose that we have h linear functions of the x's, $\zeta_1 \ldots \zeta_h$ $(h < q)$ whose coefficients obey the orthogonality relations (23.6). These h variates are then distributed independently, normally and with unit variance.

It is now possible to find $q - h$ further variates $\zeta_{h+1} \ldots \zeta_q$ which are orthogonal among themselves and to $\zeta_1 \ldots \zeta_h$. Geometrically this is evident from the possibilities of rotations in the q-way space. Algebraically it follows from the consideration that if qh of the λ's in (23.6) are known, $q(q-h)$ are unknown, and the number of conditions they must obey is

$$\tfrac{1}{2}q(q+1) - \tfrac{1}{2}h(h+1) = \tfrac{1}{2}(q-h)(q+h+1),$$

so that values of the unknowns can be found in at least one way if

$$\tfrac{1}{2}(q+h+1) \leqslant q$$

or

$$h + 1 \leqslant q.$$

Now suppose we express a sum of squares of q normal variates with unit variance, say A, as the sum of two quantities B and C; and suppose that B is distributed as the sum of squares of h independent normal variates with unit variance which are linear functions of the variates entering into A. Then we can find $q - h$ such variates independent of the first h, and C must be their sum of squares. Further, the distributions of B and C are independent. By an extension of the same argument, if

$$A = A_1 + A_2 + \ldots + A_k, \qquad \ldots \ldots (23.7)$$

A is distributed as χ^2 with ν degrees of freedom, A_1 with $\nu_1, \ldots, A_{k-1}$ with ν_{k-1}; and if the variates entering into $A_1 \ldots A_{k-1}$ are mutually independent and are linear functions of those entering into A, then A_k is distributed as χ^2 with ν_k degrees of freedom, where

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_k \qquad \ldots \ldots (23.8)$$

and A_k is independent of $A_1, \ldots A_{k-1}$.

23.6. As an extension and kind of converse of this theorem we have the result, due to Cochran, that if $A_1 \ldots A_k$ are distributed as χ^2 with $\nu_1 \ldots \nu_k$ degrees of freedom, and their sum A is distributed as χ^2 with $\nu = \Sigma(\nu_j)$ degrees, then $A_1 \ldots A_k$ are independent. We will prove this for the case $k = 2$, the more general result following in a similar way.

If the characteristic function of A_1 and A_2 is $\phi(t_1, t_2)$, we have, by hypothesis,

$$\phi(t_1, 0) = \frac{1}{(1 - 2it_1)^{\frac{1}{2}\nu_1}}$$

$$\phi(0, t_2) = \frac{1}{(1 - 2it_2)^{\frac{1}{2}\nu_2}}$$

and

$$\phi(t, t) = \frac{1}{(1 - 2it)^{\frac{1}{2}(\nu_1+\nu_2)}}.$$

Hence

$$\phi(t, t) = \phi(t, 0)\,\phi(0, t) = \frac{1}{(1 - 2it)^{\frac{1}{2}(\nu_1+\nu_2)}},$$

and thus $\phi(t, 0)$ and $\phi(0, t)$ are both divisible by a factor in $(1 - 2it)^{-1}$ and no other

factor in t because of the symmetry of $\phi(t_1, t_2)$. These factors are identified by $\phi(t_1, 0)$ and $\phi(0, t_2)$ as $(1 - 2it)^{-\frac{1}{2}\nu_1}$ and $(1 - 2it)^{-\frac{1}{2}\nu_2}$, and hence

$$\phi(t_1, t_2) = \phi(t_1, 0)\,\phi(0, t_2),$$

or A_1 and A_2 are independent.

23.7. Let us now return to the statements in (23.4). The sum $\frac{1}{v}\Sigma(x_{ij} - x_{..})^2$ is distributed as χ^2 with $\nu = N - 1$. The sum $\frac{1}{v}\Sigma(x_{ij} - x_{.j})^2$ is so distributed with $\nu_1 = N - p$. Further, the quantities $x_{ij} - x_{.j}$ may be transformed to $N - p$ independent normal variates which are linear functions of the variates entering into the first sum. It follows from **23.5** that because of the identity (23.3) the third sum $\frac{1}{v}\Sigma n_j(x_{.j} - x_{..})^2$ is distributed as χ^2 with $\nu_2 = (N - 1) - (N - p) = p - 1$ degrees of freedom, and that independently of the second sum.

Thus we may exhibit our break-up of the total sum in the following form :—

TABLE 23.1

Form of Analysis of Variance for One-way Classification.

Sum of Squares.		d.f.	Quotient.
Of family means about the mean of the whole	$\sum_j n_j(x_{.j} - x_{..})^2$	$p - 1$	$\frac{1}{p-1}\sum_j n_j(x_{.j} - x_{..})^2$
Of individuals in families about the respective family mean	$\sum_{i,j}(x_{ij} - x_{.j})^2$	$N - p$	$\frac{1}{N-p}\sum_{i,j}(x_{ij} - x_{.j})^2$
Of individuals about the mean of the whole	$\sum_{i,j}(x_{ij} - x_{..})^2$	$N - 1$	$\frac{1}{N-1}\sum_{i,j}(x_{ij} - x_{..})^2$

We note that the sums of squares and the degrees of freedom in the first two rows sum to those in the third row (though the quantities in the quotient column are not additive). This is the origin of the expression " analysis of variance," though, to be accurate, it is the sum of squares of the total which is analysed.

To avoid cumbrous phrases we refer to the sum of squares of family means about the mean of the whole as the sum of squares " between families," and to that of individuals about the respective family-means (for the time being) as " residual." We shall also speak of *total* sum of squares and *total* mean with the obvious significance, and denote degrees of freedom by the initial letters " d.f." *

23.8. Since the mean value of χ^2 with ν degrees of freedom is ν, the quotients in

* The need has been felt for a word to denote " sum of squares about the mean ". Professor Pitman has suggested the word " squariance ", though he seems to feel that this leaves something to be desired. In my own notes I use the word " deviance " but have not ventured to introduce it into the text.

(23.1) are all unbiassed estimators of v, the parent variance. Only the first two, however, are independent. We recall that the ratio

$$z = \frac{1}{2}\log \frac{N-p}{p-1}\frac{\Sigma n_j (x_{.j} - x_{..})^2}{\Sigma (x_{ij} - x_{.j})^2} \quad . \quad . \quad . \quad . \quad (23.9)$$

is distributed in Fisher's form, which is independent of the variance v. This distribution accordingly provides a convenient test of significance in the normal case.

Example 23.1

Let us consider the application of the foregoing theory to a simple example which has been chosen to reduce the arithmetic to a small amount. The following shows the lives in hours of four batches of electric lamps :—

Batch 1 : 1600, 1610, 1650, 1680, 1700, 1720, 1800.
Batch 2 : 1580, 1640, 1640, 1700, 1750.
Batch 3 : 1460, 1550, 1600, 1620, 1640, 1660, 1740, 1820.
Batch 4 : 1510, 1520, 1530, 1570, 1600, 1680.

We know that the batches were made from four different specimens of wire, but were otherwise made under identical conditions. (This, of course, over-simplifies the problem as it is encountered in practice, but will serve for purposes of illustration.) The question is, do the batches differ among themselves in length of life ? If so, we suspect that the quality of wire is varying materially, and if the lamps are to be standardised as far as possible the quality of wire must be made more uniform from batch to batch before manufacture is undertaken. The numbers in this example are small, but not much smaller than would be desirable in practice, owing to the expense and time involved in testing a lamp by running it until it burns out.

The sums of x and x^2 for the four batches will be found to be—

	Number in Sample.	$\Sigma (x)$	$\Sigma (x^2)$
Batch 1	7	11,760	19,785,400
,, 2	5	8,310	13,828,100
,, 3	8	13,090	21,503,700
,, 4	6	9,410	14,778,700
TOTALS . .	26	42,570	69,895,900

Thus for the mean life of lamp in the four batches we have 11,760/7 = 1680 ; 8310/5 = 1662 ; 13,090/8 = 1636·25 ; 9410/6 = 1568·33. These certainly differ, but is the variation such as cannot have arisen by mere sampling fluctuations ?

We find

$$x_{..} = 42{,}570/26 = 1637{\cdot}3077.$$

Thus

$$\begin{aligned}\Sigma (x_{ij} - x_{..})^2 &= \Sigma x_{ij}^2 - N x_{..}^2 \\ &= 69{,}895{,}900 - 69{,}700{,}189 \\ &= 195{,}711.\end{aligned}$$

We also have

$$\sum_j n_j (x_{.j} - x_{..})^2 = \Sigma (n_j x_{.j}) x_{.j} - N x_{..}^2$$
$$= 44{,}360.$$

The analysis then takes the form—

Sum of Squares.		d.f.	Quotient.
Between batches	44,360	3	14,787
Residual	151,351	22	6,880
TOTALS	195,711	25	7,828

We have

$$z = \tfrac{1}{2} \log_e \frac{14{,}787}{6880} = 0{\cdot}383$$
$$\nu_1 = 3, \qquad \nu_2 = 22.$$

The 5-per-cent. point for these degrees of freedom is seen from the tables to be 0·5574. The observed value is therefore not significant, and we conclude that, so far as this test is concerned, there is nothing to throw doubt on the homogeneity of the group.

Having decided, provisionally at least, to accept the hypothesis that the data are homogeneous, we may ask, what is the best estimate of the parent variance ? Our analysis has given three different estimates, viz. 14,787, 6880 and 7838. It seems natural to use the last, which depends on the greatest number of degrees of freedom.

With this value we find for the variance of the mean of samples of n,

$$\sqrt{\frac{7828}{n}} = \frac{88{\cdot}48}{\sqrt{n}}.$$

The greatest difference of means observed is that between the first and fourth batch, $1680 - 1568{\cdot}33 = 111{\cdot}67$. The standard error of this difference is

$$88{\cdot}48 \sqrt{\left(\tfrac{1}{7} + \tfrac{1}{6}\right)} = 49{\cdot}2.$$

The observed difference is rather more than twice the standard error, but we cannot conclude that it is significant on that account. In fact, we have picked out the *greatest* difference for examination from the six possible comparisons of pairs, and the distribution of the greatest difference must have a larger standard error than that of a difference chosen at random, which is what we have found. Nevertheless the fact that even the greatest difference is only slightly in excess of twice the standard error affords some general evidence in support of the hypothesis of homogeneity.

We may also note that if a more accurate test of the difference of two means is required the t-test may be invoked ; but here also we must remember that we are testing the greatest of a set of differences. Where there are only two families concerned, the analysis of variance reduces to the t-test for the difference of sample means when variances of the parents are assumed equal.

23.9. Suppose now that in the case of one classification we have applied a test by means of the analysis of variance and have found that the hypothesis of homogeneity is

unacceptable, or, in plain English, that the parents do differ. Let us then consider the alternative that the populations are still normal and that they differ in their means *but not in their variances.*

At first sight this may seem a highly artificial assumption to make, for if the populations differ in their means it is not unlikely that they may differ in other respects. This is undoubtedly so, but if there is serious possibility of difference in variances their homogeneity may be discussed separately by means of tests we shall consider in Chapter 26. Apart from this, there often arise in practice situations in which approximate equality of variance is plausible on prior grounds. For instance, we may be testing the effect of manuring on cereal yields, and it is reasonable to suppose that if the manure exerts any effect at all it will increase all plants of the same variety to about the same extent—that it will, in fact, displace the location of the distribution of yields without affecting its dispersion.

23.10. The question we have now to consider is whether we can make an estimate of the common variance of the populations. A little thought will show that we can. The reasoning which led to the conclusion that the *residual* sum of squares is distributed as $v\chi^2$ with $N-p$ degrees of freedom remains unchanged, so that the residual quotient in Table 23.1 continues to provide an estimator of v. The other two no longer do so. Consider, in fact, the sum of squares between families, and let the mean of the jth family be $m_{.j}$. Then we have

$$\begin{aligned} E\sum_j n_j(x_{.j}-x_{..})^2 &= E\sum_j n_j\{x_{.j}-m_{.j}-(x_{..}-m_{..})+m_{.j}-m_{..}\}^2 \\ &= E\sum_j n_j\{x_{.j}-m_{.j}-(x_{..}-m_{..})\}^2+\sum_j n_j(m_{.j}-m_{..})^2. \qquad (23.10) \end{aligned}$$

Here $m_{..}$ is the mean $\frac{1}{N}\sum_j n_j m_{.j}$ and hence $x_{.j}-m_{.j}$ has the mean $x_{..}-m_{..}$. Thus $\sum n_j\{x_{.j}-m_{.j}-(x_{..}-m_{..})\}^2$ is distributed as $v\chi^2$ with $p-1$ degrees of freedom and

$$E\sum_j n_j(x_{.j}-x_{..})^2=(p-1)v+\sum n_j(m_{.j}-m_{..})^2. \qquad (23.11)$$

Not unless $m_{.j}=m_{..}$—that is, all populations have the same mean—does the expression on the right reduce to $(p-1)v$, and hence the quotient between families give an unbiassed estimator of v. In other cases it is greater.

Similarly,

$$\begin{aligned} E\sum_{i,j}(x_{ij}-x_{..})^2 &= E\sum_{i,j}\{x_{ij}-m_{.j}-(x_{..}-m_{..})\}^2+E\sum_{i,j}(m_{.j}-m_{..})^2 \\ &= (N-1)v+\sum_j n_j(m_{.j}-m_{..})^2 \qquad (23.12) \end{aligned}$$

The expectation of the difference of the two terms considered in (23.11) and (23.12) confirms that the residual sum of squares provides an estimator of $(N-p)v$.

23.11. A comparison of the formulae we have already reached and those of section **14.31** will show that the study of intra-class correlation is very closely related to the analysis of variance. It is an interesting exercise to derive the z-test directly from the sampling distribution of intra-class r given in equation (14.110) (vol. I, p. 362) and vice-versa.

Two-way Classification

23.12. We proceed to the case when the variate-values belong not to one of a single set of families but to two, say A and B. In the first instance we shall consider the situation

when there is only a single value in the jth class of A and the kth class of B. Our sample may then be set out in the tabular form :

		CLASS B					
		B_1	B_2	B_3	. . .	B_q	TOTALS
CLASS A	A_1	x_{11}	x_{12}	x_{13}	. . .	x_{1q}	$qx_{1.}$
	A_2	x_{21}	x_{22}	x_{23}	. . .	x_{2q}	$qx_{2.}$
	A_3	x_{31}	x_{32}	x_{33}	. . .	x_{3q}	$qx_{3.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$ $\vdots$ $\vdots$	$\vdots$	$\vdots$
	A_p	x_{p1}	x_{p2}	x_{p3}	. . .	x_{pq}	$qx_{p.}$
	TOTALS	$px_{.1}$	$px_{.2}$	$px_{.3}$	. . .	$px_{.q}$	$pqx_{..}$

. . . (23.13)

This is not a contingency table. The numbers x_{jk} are variate-values, not frequencies. As usual, $x_{j.}$ signifies the mean of values in the class A_j and $x_{.k}$ the mean of values in the class B_k, $x_{..}$ being the mean of the whole.

We have the algebraic identity

$$\begin{aligned}\sum_{j,k}(x_{jk}-x_{..})^2 &= \sum_{j,k}(x_{jk}-x_{j.}-x_{.k}+x_{..}+x_{j.}-x_{..}+x_{.k}-x_{..})^2\\ &= \sum_{j,k}(x_{jk}-x_{j.}-x_{.k}+x_{..})^2+\sum_{j,k}(x_{j.}-x_{..})^2+\sum_{j,k}(x_{.k}-x_{..})^2\\ &= \sum_{j,k}(x_{jk}-x_{j.}-x_{.k}+x_{..})^2+q\sum_{j}(x_{j.}-x_{..})^2+p\sum_{k}(x_{.k}-x_{..})^2 \qquad (23.14)\end{aligned}$$

the cross-product terms vanishing on summation in the usual way.

23.13. We are interested in the variation of the x's according to class membership. Let us take as our hypothesis that the pq values are homogeneous, that is to say that they all emanate from (normal) populations with the same mean m and variance v. In such a case class-membership exerts no influence on variate-values, and the observed differences are pure sampling effects.

The expression on the left in (23.14) is then distributed as $v\chi^2$ with $pq - 1$ degrees of freedom. The mean $x_{j.}$ is distributed normally with variance v/q and thus $\sum_j q\,(x_{j.} - x_{..})^2$ is distributed as $v\chi^2$ with $p - 1$ d.f. Similarly, $\sum_k p\,(x_{.k} - x_{..})^2$ is so distributed with $q - 1$ d.f. Finally the remaining term on the right is distributed as $v\chi^2$ with $(p - 1)(q - 1)$ d.f. ; for each term is normal with variance $\dfrac{(p-1)(q-1)}{pq}v$, since

$$\begin{aligned}x_{jk}-x_{j.}-x_{.k}+x_{..} &= x_{jk}\left(1-\frac{1}{q}-\frac{1}{p}+\frac{1}{pq}\right)-\sum_{l}x_{jl}\left(\frac{1}{q}-\frac{1}{pq}\right)\\ &\quad -\sum_{m}x_{mk}\left(\frac{1}{p}-\frac{1}{pq}\right)+\frac{1}{pq}\sum_{l,m}x_{lm}, \qquad l\neq j,\ m\neq k\end{aligned}$$

so that the sum of squares of coefficients on the right is

$$\left\{\frac{(p-1)(q-1)}{pq}\right\}^2 + (q-1)\left(\frac{p-1}{pq}\right)^2 + (p-1)\left(\frac{q-1}{pq}\right)^2 + \frac{(p-1)(q-1)}{(pq)^2}$$
$$= \frac{(p-1)(q-1)}{pq}. \quad . \quad . \quad . \quad . \quad . \quad (23.15)$$

Thus, since there are $p + q - 1$ linear relations connecting the pq quantities

$$x_{jk} - x_{j.} - x_{.k} + x_{..},$$

their sum of squares is distributed as $v\chi^2$ with $pq - (p + q - 1) = (p - 1)(q - 1)$ degrees of freedom, which checks against the mean value of the individual square given by (23.15).

We may thus analyse the variance in the following way :—

TABLE 23.2

Form of Analysis of Variance for Two-way Classification with One Member in each Subclass

	Sums of Squares.	d.f.	Quotient.
Between A-classes	$q \sum_j (x_{j.} - x_{..})^2$	$p - 1$	$\frac{q}{p-1} \sum_j (x_{j.} - x_{..})^2$
Between B-classes	$p \sum_k (x_{.k} - x_{..})^2$	$q - 1$	$\frac{p}{q-1} \sum_k (x_{.k} - x_{..})^2$
Residual . . .	$\sum_{j,k} (x_{jk} - x_{j.} - x_{.k} + x_{..})^2$	$(p-1)(q-1)$	$\frac{1}{(p-1)(q-1)} \sum_{j,k} (x_{jk} - x_{j.} - x_{.k} + x_{..})^2$
TOTALS . .	$\sum_{j,k} (x_{jk} - x_{..})^2$	$pq - 1$	

The sums of squares and degrees of freedom (but not the quotients) are additive as before. It follows from the theorem of **23.6** that the three constituent sums are independent. Each quotient provides an unbiassed estimator of v.

23.14. Our use of these results proceeds by an easy generalisation of the method exemplified in Example 23.1. We take as our hypothesis the supposition that all samples are from normal populations with identical mean and variance. Comparison of the estimates in the quotient column then provides a test of significance. If the hypothesis is rejected we may examine the alternative that means are different but variances identical throughout, in which case we shall find that the residual still provides an estimate of the variance, provided that an important additional assumption is made.

Example 23.2

The following data (Daniels, *Supp. J.R.S.S.*, 1938, **5**, 89) show the weight in grams of 95-yard lengths of wool thread from 100 " ends " being spun on four bobbins, 25 ends

to the bobbin. We are interested in two factors, the variation between bobbins and the variation in the 25 ends on the same bobbin, according to their position.

TABLE 23.3

Weight in Grams of 100 95-yard Lengths of Wool Thread spun on Four Bobbins.

End Number.	Bobbin Number. 1	2	3	4	Totals.
1	7·50	7·23	7·50	7·53	29·76
2	7·52	7·81	7·77	8·05	31·15
3	7·70	7·94	7·83	8·16	31·63
4	7·93	7·94	7·96	7·76	31·59
5	7·78	7·89	8·02	7·85	31·54
6	7·73	8·23	7·99	8·14	32·09
7	8·07	8·27	8·25	8·26	32·85
8	8·01	8·54	8·24	8·54	33·33
9	8·22	8·24	8·37	8·10	32·93
10	8·24	8·35	8·43	8·15	33·17
11	8·17	8·29	8·46	8·38	33·30
12	8·09	8·54	8·33	8·47	33·43
13	8·11	8·45	8·27	8·38	33·21
14	7·96	8·43	8·24	8·60	33·23
15	8·09	8·47	8·12	8·45	33·13
16	8·04	8·33	8·14	8·43	32·94
17	7·78	8·47	8·19	8·57	33·01
18	8·11	8·63	8·36	8·38	33·48
19	8·17	8·31	8·31	8·16	32·95
20	8·12	8·31	8·47	8·41	33·31
21	8·13	8·10	8·19	8·27	32·69
22	8·01	8·01	8·37	7·96	32·35
23	8·17	7·92	8·27	8·08	32·44
24	8·05	8·27	8·07	8·16	32·55
25	7·91	7·92	8·28	8·52	32·63
Totals .	199·61	204·89	204·43	205·76	814·69

It simplifies the arithmetic if we take a working mean at 8·00. The total sum of squares about this mean is then found to be

$$\Sigma\,(x_{jk})^2 = 9{\cdot}3829,$$

and we have also

$$\Sigma\,(x_{jk}) = 14{\cdot}69.$$

Hence

$$\Sigma\,(x_{jk} - x_{..})^2 = 9{\cdot}3829 - (0{\cdot}1469)\,(14{\cdot}69)$$
$$= 7{\cdot}224{,}939.$$

The means of the four bobbins are

$$7{\cdot}9844,\quad 8{\cdot}1956,\quad 8{\cdot}1772,\quad 8{\cdot}2304.$$

With the same working mean we find for the sum of squares

$$\Sigma\,(x_{.k})^2 = 0{\cdot}122{,}986{,}72\ ;$$

and hence

$$p\,\Sigma (x_{.k} - x_{..})^2 = 25\,(0{\cdot}122{,}986{,}72) - (0{\cdot}1469)\,(14{\cdot}69)$$
$$= 0{\cdot}916{,}707.$$

The means of the four ends of corresponding position on the four bobbins can, of course, be found from the totals in the last column of the table, but it is simpler to find $\Sigma (qx_{j.} - qx_{..})^2$ and then divide by q^2. We find

$$\Sigma (x_{j.} - x_{..})^2 = \frac{4\,(27{\cdot}1831)}{16} - (0{\cdot}1469)\,(14{\cdot}69)$$
$$= 4{\cdot}637{,}814.$$

The continual appearance of the factor $(0{\cdot}1469)\,(14{\cdot}69) = Nx^2_{..}$ is to be noted. The quantity is best computed once for all at the outset.

The residual sum of squares is then obtainable by subtraction, and we have the following analysis :—

TABLE 23.4

Analysis of Variance for the Data of Table 23.3.

Sums of Squares.		d.f.	Quotient.
Between bobbins	0·916,707	3	0·3056
Between ends	4·637,814	24	0·1932
Residual	1·670,418	72	0·0232
TOTALS	7·224,939	99	0·0730

The variation between bobbins and that between ends are both significant—the ratio of the corresponding quotients to the residual quotient is so big in each case as hardly to require the z-test. We are led to suspect that the variation between bobbins, small as it is, cannot be a chance effect, and it looks as if bobbin number 1 is not getting its fair share of thread. Similarly, the weight of thread seems to be dependent on whereabouts the thread is spun on the bobbins, and an inspection of the original data suggests a systematic variation as we proceed along the bobbin from end number 1 to end number 25, with a possible maximum in the middle. If the manufacturing process is to be standardised as much as possible, we should have to examine the reasons for the shortage of weight on the first bobbin and for this systematic effect of position on the bobbin.

23.15. Suppose now that, as in the example just given, the hypothesis of homogeneity is rejected. What interpretation can we put on the residual quotient ? Let us assume that each observation comes from a normal population with variance v, but that the parent mean of the subclass $A_j\,B_k$ is m_{jk}, these quantities varying from one subclass to another. Is the residual quotient an unbiassed estimator of v ? In general the answer is " no ", but there is an important class of case in which it is affirmative.

Let $m_{j.}$ be the mean of the q values of m_{jk} in the class A_j, $m_{.k}$ that of the p values in B_k, and $m_{..}$ the mean of the whole set of m's. Then we may write

$$x_{jk} = m_{jk} + \xi_{jk} \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.16)$$
$$x_{j.} = m_{j.} + \xi_{j.},\ \text{etc.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.17)$$

Then

$$E\,\Sigma\,(x_{jk}-x_{j.}-x_{.k}+x_{..})^2=E\,\Sigma\,(m_{jk}-m_{j.}-m_{.k}+m_{..}+\xi_{jk}-\xi_{j.}-\xi_{.k}+\xi_{..})^2$$
$$=E\,\Sigma\,(m_{jk}-m_{j.}-m_{.k}+m_{..})^2+E\,\Sigma\,(\xi_{jk}-\xi_{j.}-\xi_{.k}+\xi_{..})^2, \quad . \quad (23.18)$$

the product term vanishing as usual. The second term on the right is equal to $(p-1)(q-1)v$, for the ξ's are distributed with variance v about *zero* mean, so that the term in question is the residual sum of squares in a $p \times q$ two-way classification of a homogeneous sample and hence has the stated expectation. Thus we have

$$E\,\Sigma\,(x_{jk}-x_{j.}-x_{.k}+x_{..})^2 = \Sigma(m_{jk}-m_{j.}-m_{.k}+m_{..})^2+(p-1)(q-1)v. \quad (23.19)$$

The residual quotient will then provide an unbiassed estimator of v if and only if

$$m_{jk}-m_{j.}-m_{.k}+m_{..}=0. \quad . \quad . \quad . \quad . \quad (23.20)$$

23.16. Now suppose that x_{jk} is made up of three parts which are additive, viz.

(1) the effect of the class A_j, say a_j;

(2) the effect of the class B_k, say b_k; and

(3) a residual ζ_{jk} which is normal and has zero mean.

This kind of hypothesis will recur frequently. It amounts to an assumption that there is in x_{jk} an element a_j which affects alike all members of the class A_j but varies from one A-class to another; an element b_k which similarly affects alike all members of B_k but varies from B-class to B-class; and a third component representing random variation which, apart from the sampling factor, is the same for all subclasses $A_j\,B_k$. We then have

$$x_{jk}=a_j+b_k+\zeta_{jk} \quad . \quad . \quad . \quad . \quad . \quad (23.21)$$

and

$$\left.\begin{aligned} m_{jk} &= a_j + b_k \\ m_{j.} &= a_j + b_. \\ m_{.k} &= a_. + b_k \\ m_{..} &= a_. + b_. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.22)$$

where, as usual, the subscript periods in the a's and b's denote averaging. Thus

$$m_{jk}-m_{j.}-m_{.k}+m_{..}=a_j+b_k-(a_j+b_.)-(a_.+b_k)+a_.+b_.$$
$$=0,$$

so that (23.20) is satisfied and the residual quotient is an unbiassed estimator of the variance v.

Under the same conditions it will be found that

$$q\,E\,\sum_j\,(x_{j.}-x_{..})^2=(p-1)\,v+q\sum_j\,(m_{j.}-m_{..})^2$$
$$=(p-1)\,v+q\sum_j\,(a_j-a_.)^2 \quad . \quad . \quad . \quad . \quad (23.23)$$

$$p\,E\,\sum_k\,(x_{.k}-x_{..})^2=(q-1)\,v+p\sum_k\,(b_k-b_.)^2 \quad . \quad . \quad . \quad . \quad (23.24)$$

$$E\,\Sigma\,(x_{jk}-x_{..})^2=(pq-1)\,v+\sum_{j,\,k}\,(a_j-a_.+b_k-b_.)^2$$
$$=(pq-1)\,v+q\sum_j\,(a_j-a_.)^2+p\sum_k\,(b_k-b_.)^2 \quad . \quad (23.25)$$

23.17. We have supposed that the component ζ had a zero mean, but of course if all these components had the same mean, the constant common to them could be absorbed

into the functions a_j and b_k. Our hypothesis is thus a little more general than it appears. In certain practical cases it is a plausible hypothesis to make. For instance, in Example 23.2 it is reasonable to suppose that the effect of a particular bobbin is the same for all ends, and the effect of situation the same for all bobbins. If there is any serious doubt on the point we have to collect further data and consider interactions in the manner described later (see **23.22**).

It may, however, be noted that if the variation of the m_{jk}'s is comparatively small the appearance of the term containing them in (23.19) does not materially vitiate an estimate of v from the residual quotient. In any case that estimate will be greater than the unbiassed estimate, so that our inferences about significant differences of mean values will, properly interpreted, be on the safe side.

23.18. Before going farther we may remark that the quantity we have called the residual sum of squares and the associated quotient are often referred to as "error" or "interaction" terms. The former is likely to cause misunderstanding and is better avoided altogether, for, as we have seen, it provides a measure of sampling variance, and therefore of experimental error, only in particular cases. The word "interaction" we shall define below; it has been used in different senses by different writers, and when consulting original memoirs the reader should endeavour to ascertain the precise meaning which is being attached to it—if he can. In considering a given analysis it is as well to reflect on the precise nature of the items covered by such expressions as "residual", "remainder", "error" and so forth.

Three-way Classification

23.19. Consider now the case when there are three classifications into A-, B- and C-classes. As before, we shall consider in the first place one member in each subclass $A_j B_k C_l$, typified by x_{jkl}. We now have

$$\begin{aligned}\sum_{j,k,l}(x_{jkl}-x_{\dots})^2 &= \Sigma(x_{j..}-x_{\dots})^2+\Sigma(x_{.k.}-x_{\dots})^2+\Sigma(x_{..l}-x_{\dots})^2\\ &\quad+\Sigma(x_{jk.}-x_{j..}-x_{.k.}+x_{\dots})^2+\Sigma(x_{j.l}-x_{j..}-x_{..l}+x_{\dots})^2\\ &\quad+\Sigma(x_{.kl}-x_{.k.}-x_{..l}+x_{\dots})^2\\ &\quad+\Sigma(x_{jkl}-x_{jk.}-x_{j.l}-x_{.kl}+x_{j..}+x_{.k.}+x_{..l}-x_{\dots})^2, \quad (23.26)\end{aligned}$$

the summations extending over all members of the sample, pqr in number, so that we may replace expressions such as $\sum_{j,k,l}(x_{j..}-x_{\dots})^2$ by $qr\sum_j(x_{j..}-x_{\dots})^2$, etc.

On the usual hypothesis of normality and homogeneity we find that the first three terms on the right of (23.26) are distributed as $v\chi^2$ with $p-1$, $q-1$ and $r-1$ degrees of freedom. The second group is so distributed with $(p-1)(q-1)$, $(p-1)(r-1)$ and $(q-1)(r-1)$ degrees of freedom. The last is distributed with $(p-1)(q-1)(r-1)$ degrees of freedom. All but the last of these results follow from the two-way case, and the last may be established as in **23.13** or by the consideration that for any fixed l the term has $(p-1)(q-1)$ degrees of freedom and that there are $(r-1)$ independent l's.

We may then write the analysis in the form shown in Table 23.5. (For the present the expression "interaction AB" is to be regarded merely as a name given to a particular sum of squares. As before, the sums of squares and degrees of freedom are additive,

and the seven items into which the total sum of squares is analysed are distributed independently.)

TABLE 23.5

Form of Analysis of Variance for Three-way Classification with One Member in each Subclass.

Sum of Squares.		d.f.	Quotient.
Between A-classes	$\Sigma(x_{j..} - x_{...})^2$	$p - 1$	The quotient of the sum of squares by the corresponding d.f.
Between B-classes	$\Sigma(x_{.k.} - x_{...})^2$	$q - 1$	
Between C-classes	$\Sigma(x_{..l} - x_{...})^2$	$r - 1$	
Interaction AB	$\Sigma(x_{jk.} - x_{j..} - x_{.k.} + x_{...})^2$	$(p-1)(q-1)$	
Interaction BC	$\Sigma(x_{.kl} - x_{.k.} - x_{..l} + x_{...})^2$	$(q-1)(r-1)$	
Interaction CA	$\Sigma(x_{j.l} - x_{j..} - x_{..l} + x_{...})^2$	$(r-1)(p-1)$	
Residual	$\Sigma(x_{jkl} - x_{j..} - x_{.k.} - x_{..l} + x_{jk.} + x_{.kl} + x_{j.l} - x_{...})^2$	$(p-1)(q-1)(r-1)$	
TOTALS	$\Sigma(x_{jkl} - x_{...})^2$	$pqr - 1$	

23.20. If the hypothesis of homogeneity is rejected we may consider the alternative represented by

$$x_{jkl} = a_j + b_k + c_l + \zeta_{jkl}, \qquad (23.27)$$

where ζ, as usual, is normal with zero mean. As in **23.16** it will be found that the residual term in Table 23.5 has expectation $(p-1)(q-1)(r-1)v$, and hence continues to provide an unbiassed estimator of v. The quotients between classes are affected like those in equations (23.23) to (23.25); but the interaction terms also provide estimators of v with the appropriate degrees of freedom. For instance,

$$\begin{aligned}(x_{jk.} - x_{j..} - x_{.k.} + x_{...}) &= a_j + b_k + c_. + \zeta_{jk.} - (a_j + b_. + c_. + \zeta_{j..}) \\ &\quad - (a_. + b_k + c_. + \zeta_{.k.}) + (a_. + b_. + c_. + \zeta_{...}) \\ &= \zeta_{jk.} - \zeta_{j..} - \zeta_{.k.} + \zeta_{...} \qquad (23.28)\end{aligned}$$

so that the expectation of the sum of squares of the x-terms is that of the ζ-terms, which we know to be $(p-1)(q-1)v$.

23.21. This brings up a new point arising for the first time in the three-way classification. If (23.27) is true, the analysis of variance will provide four different estimators of the variance v, namely the interactions AB, BC and CA and the residual. These are independent (for they depend only on the ζ's, and the theory appropriate to the case of homogeneity continues to apply) and their ratios may be tested in the z-distribution. If these ratios are such as can have arisen from random sampling we may accept the hypothesis represented by (23.27); if not we must reject it. In short, the interaction quotients provide a test of the hypothesis (23.27). In the two-way classification no such test is available.

Interactions

23.22. On the hypothesis (23.27) the interaction quotients of type AB give unbiassed estimators of the variance v. If in any particular case these quotients differ significantly among themselves or from any other independent estimator of v, we have to reject the hypothesis. Apart from the normality of the variation of ζ, which is not for the moment in question, this means that we cannot represent the data as the sum of separate effects due to A-, B- and C-classes, together with a residual ζ which is the same in form for all

subclasses. The effects of the classes are entangled—or, as we may say, they *interact.* This is the origin of the term "interaction".

Suppose, for instance, our data are crop-yields, and membership of the three classes corresponds to applications of three manures, nitrogen (A), potash (B) and phosphate (C). The hypothesis represented by (23.27) would then be equivalent to supposing that all three manures exerted an effect on yields, but that they did so independently. A given dressing of nitrogen would increase the yield by a_j, whatever dressings of the other fertilisers were applied. But it might happen that the response in yield to a_j varied according to how much of the others were present—potash might either stimulate the effect of nitrogen or inhibit it. If this were so, the fertilisers would interact and the hypothesis (23.27) would break down. Significant departures from homogeneity in the interaction terms usually lead us to search for possible entanglements of this kind.

23.23. It must not be overlooked, however, that significant interactions do not necessarily imply interaction in any real sense. They may arise from heterogeneity in the data. To return to our example of crop-yields, suppose the yields were taken from a series of plots which differed materially in natural fertility. It might very well be found that the hypothesis (23.27) could not be justified even if the differences in yields due to the natural effect were partially absorbed into the coefficients a, b and c. If by chance the heavier dressings of fertilisers were applied to plots of greater fertility, the hypothesis might be shown as failing and "significant" interactions appear. Such points as this require careful consideration in the interpretation of significance, and we shall illustrate them in some examples below.

23.24. Interactions of type AB, involving two classes, are said to be of the first order. When considering the general n-way classification we shall see that there can appear interactions of second, third, fourth . . . order. In fact, the residual in Table 23.5 is formally equivalent to an interaction of the second order, of type ABC, just as the first-order interaction is equivalent to the residual in the two-way analysis of Table 23.2.

To complete the definitions, we may define the sum of squares between A-classes as an interaction of order zero. The seven constituent items in Table 23.5 would then correspond to the following :—

	Interaction.	d.f.
Order zero	A	$p-1$
	B	$q-1$
	C	$r-1$
Order 1	AB	$(p-1)(q-1)$
	BC	$(q-1)(r-1)$
	CA	$(r-1)(p-1)$
Order 2	ABC	$(p-1)(q-1)(r-1)$

This illustrates the general symmetry of the analysis and suggests obvious generalisations.

n-way Classifications

23.25. For instance, with five classes A, B, C, D and E we may analyse the total sums of squares into $2^5 - 1 = 31$ components. There will be $\binom{5}{1} = 5$ interactions of

order zero; $\binom{5}{2} = 10$ interactions of first order, type AB; $\binom{5}{3} = 10$ interactions of second order, type ABC; $\binom{5}{4} = 5$ interactions of third order, type $ABCD$; and one residual or interaction of fourth order, type $ABCDE$. The interactions of zero, first and second order are of a type already familiar :—

$$\begin{aligned} &\Sigma (x_{j....} - x_{.....})^2 \\ &\Sigma (x_{jk...} - x_{j....} - x_{.k...} + x_{.....})^2 \\ &\Sigma (x_{jkl..} - x_{jk...} - x_{.kl..} - x_{j.l..} + x_{j...} + x_{.k...} + x_{..l.} - x_{.....})^2 \end{aligned} \quad . \quad (23.29)$$

The third-order interactions are typified by

$$\begin{aligned} \Sigma (x_{jklm.} - x_{jkl..} - x_{.klm.} - x_{j.lm.} - x_{jk.m.} + x_{jk...} + x_{j.l..} + x_{j..m.} \\ + x_{.kl..} + x_{.k.m.} + x_{..lm.} - x_{j....} - x_{.k...} - x_{..l..} - x_{...m.} + x_{.....})^2 \end{aligned} \quad . \quad (23.30)$$

and the reader will be able to write down the residual for himself.

As usual, the 31 terms all furnish independent estimators of the variance on the hypothesis of homogeneity, and if this is rejected we may consider the alternative represented by

$$x_{jklmn} = a_j + b_k + c_l + d_m + e_n + \zeta_{jklmn} \quad . \quad . \quad . \quad . \quad (23.31)$$

The complete analysis in such cases may become very complex, but frequently it is sufficient to consider only sums of squares suggested for investigation by prior expectations.

Example 23.3

The following data show the percentage water-content in a number of samples of a commercial product. Six samples were chosen; each sample was tested by four different operators; and each operator carried out the determination by three different methods. We have thus a $6 \times 4 \times 3$ classification.

TABLE 23.6

Percentage Water-Content of Six Samples determined by Four Operators using Three Methods.

Samples.	Operators.											
	1			2			3			4		
	Tests.			Tests.			Tests.			Tests.		
	1	2	3	1	2	3	1	2	3	1	2	3
1	59	61	61	57	60	58	55	58	62	54	56	59
2	57	58	60	57	58	58	61	60	57	60	56	58
3	55	57	59	55	55	56	54	52	58	53	55	55
4	60	57	58	56	57	57	54	58	55	61	59	58
5	61	61	60	59	58	59	61	57	60	62	60	60
6	63	59	60	62	63	61	64	62	59	59	60	61

We will first of all analyse the variance systematically with rather more arithmetical detail than is usually required, in order to illustrate the process.

A great deal of work is saved if we take a mean at 60. The table then becomes—

TABLE 23.7

Samples.	Operators.																Totals
	1				2				3				4				
	Tests.				Tests.				Tests.				Tests.				
	1	2	3	Totals	1	2	3	Totals	1	2	3	Totals	1	2	3	Totals	
1	−1	1	1	1	−3	0	−2	−5	−5	−2	2	−5	−6	−4	−1	−11	−20
2	−3	−2	0	−5	−3	−2	−2	−7	1	0	−3	−2	0	−4	−2	−6	−20
3	−5	−3	−1	−9	−5	−5	−4	−14	−6	−8	−2	−16	−7	−5	−5	−17	−56
4	0	−3	−2	−5	−4	−3	−3	−10	−6	−2	−5	−13	1	−1	−2	−2	−30
5	1	1	0	2	−1	−2	−1	−4	1	−3	0	−2	2	0	0	2	−2
6	3	−1	0	2	2	3	1	6	4	2	−1	5	−1	0	1	0	13
Totals	−5	−7	−2	−14	−14	−9	−11	−34	−11	−13	−9	−33	−11	−14	−9	−34	−115

We have shown the totals of the tests for each operator, of the tests for all operators, and of samples for each test.

We now form three two-way tables from this by adding the values of one of the variates, e.g.—

TABLE 23.8

Operators.

		1	2	3	4	Totals.
Samples.	1	1	− 5	− 5	− 11	− 20
	2	− 5	− 7	− 2	− 6	− 20
	3	− 9	− 14	− 16	− 17	− 56
	4	− 5	− 10	− 13	− 2	− 30
	5	2	− 4	− 2	2	− 2
	6	2	6	5	0	13
	Totals	− 14	− 34	− 33	− 34	− 115

TABLE 23.9

Tests.

		1	2	3	Totals.
Samples.	1	− 15	− 5	0	− 20
	2	− 5	− 8	− 7	− 20
	3	− 23	− 21	− 12	− 56
	4	− 9	− 9	− 12	− 30
	5	3	− 4	− 1	− 2
	6	8	4	1	13
	Totals	− 41	− 43	− 31	− 115

TABLE 23.10

Operators.

		1	2	3	4	Totals.
Tests.	1	− 5	− 14	− 11	− 11	− 41
	2	− 7	− 9	− 13	− 14	− 43
	3	− 2	− 11	− 9	− 9	− 31
	Totals	− 14	− 34	− 33	− 34	− 115

As we have inserted the totals of various kinds in Table 23.7 these subsidiary tables can be picked out at once ; but in general, totals are not available in the original (and for four-way classifications it is difficult to find a form of tabular presentation which will permit of their insertion) so that the tables have to be separately compiled. In practice I find it convenient to do so in any case to avoid picking out the wrong figures in the original table.

Pursuing the condensation process, we should now derive three one-way tables from Tables 23.8 to 23.10, but in fact the row and column totals already give us what is required (and incidentally provide a check on the arithmetic).

Now we proceed to find the various sums of squares. For the total of all observations we find − 115, and for the sum of squares of observations 653. Thus

$$x_{\cdots} = \frac{-115}{72} = -1{\cdot}597{,}222$$

$$Nx^2_{\cdots} = -115x_{\cdots} = 183{\cdot}680{,}556$$

$$\begin{aligned} \Sigma\,(x_{jkl} - x_{\cdots})^2 &= \Sigma\,(x_{jkl})^2 - Nx^2_{\cdots} \\ &= 653 - 183{\cdot}680{,}556 \\ &= 469{\cdot}319{,}444 \qquad (23.32) \end{aligned}$$

with $6 \times 4 \times 3 - 1 = 71$ degrees of freedom.

For the interactions of order zero we require the sums of type

$$\Sigma\,(x_{j..} - x_{...})^2 = \Sigma\,(x_{j..})^2 - Nx^2_{...},$$

where summation takes place over the N values. It is, however, unnecessary to work out the means $x_{j..}$. Consider, for example, the sum of squares between samples. From the totals of Table 22.8 or Table 22.9 we find (j denoting samples)—

$$\begin{aligned}\Sigma\,(12x_{j..})^2 &= (-20)^2 + (-20)^2 + \ldots + 13^2\\ &= 5009,\end{aligned}$$

where the summation is over six values only. Thus, for summation over the 72 values—

$$\Sigma\,(x_{j..})^2 = \frac{12}{12^2}\,5009 = 417{\cdot}416{,}667.$$

Hence

$$\begin{aligned}\Sigma\,(x_{j..} - x_{...})^2 &= 417{\cdot}416{,}667 - 183{\cdot}680{,}556\\ &= 233{\cdot}736{,}111 \quad . \quad . \quad . \quad . \quad . \quad . \qquad (23.33)\end{aligned}$$

with $6 - 1 = 5$ d.f.

Similarly (k denoting operators) we find—

$$\begin{aligned}\Sigma\,(x_{.k.} - x_{...})^2 &= \frac{3597}{18} - 183{\cdot}680{,}556\\ &= 16{\cdot}152{,}778 \quad . \quad . \quad . \quad . \quad . \qquad (23.34)\end{aligned}$$

with 3 d.f.; and (l denoting tests)—

$$\begin{aligned}\Sigma\,(x_{..l} - x_{...})^2 &= \frac{4491}{24} - 183{\cdot}680{,}556\\ &= 3{\cdot}444{,}444 \quad . \quad . \quad . \quad . \quad . \qquad (23.35)\end{aligned}$$

with two degrees of freedom.

Now we require first-order interactions. We have (summation being over the N values)—

$$\begin{aligned}\Sigma\,(x_{jk.} - x_{j..} - x_{.k.} + x_{...})^2 &= \Sigma\,(x_{jk.} - x_{...})^2 + \Sigma\,(x_{j..} - x_{...})^2\\ &\quad + \Sigma\,(x_{.k.} - x_{...})^2 - 2\Sigma\,(x_{jk.} - x_{...})\,(x_{j..} - x_{...})\\ &\quad - 2\,\Sigma\,(x_{jk.} - x_{...})\,(x_{.k.} - x_{...})\\ &= \Sigma\,(x_{jk.} - x_{...})^2 - \Sigma\,(x_{j..} - x_{...})^2 - \Sigma\,(x_{.k.} - x_{...})^2 \qquad (23.36)\end{aligned}$$

and thus the first-order interaction term is ascertainable from $\Sigma\,(x_{jk.})^2$ and quantities which have already been computed.

From the body of Table 23.8 (remembering that summation relates to 72 values and hence that each value in the table is counted 3 times) we find

$$\begin{aligned}\Sigma\,(x_{jk.})^2 &= \frac{3}{3^2}\,\{1^2 + (-5)^2 + \ldots\} = \frac{1499}{3}\\ &= 499{\cdot}666{,}667.\end{aligned}$$

The interaction term is then

$$499{\cdot}666{,}667 - 183{\cdot}680{,}556 - 233{\cdot}736{,}111 - 16{\cdot}152{,}778 = 66{\cdot}097{,}222 \quad . \qquad (23.37)$$

with $(6 - 1)\,(4 - 1) = 15$ d.f.

Similarly in the body of Table 23.9 we find for the sum of squares 1915. Hence the interaction of samples and tests is

$$\frac{1915}{4} - 183{\cdot}680{,}556 - 233{\cdot}736{,}111 - 3{\cdot}444{,}444 = 57{\cdot}888{,}889. \quad . \qquad (23.38)$$

In the body of Table 23.10 the sum of squares is 1245. Hence the interaction of tests and operators is

$$\frac{1245}{6} - 183{\cdot}680{,}556 - 16{\cdot}152{,}778 - 3{\cdot}444{,}444 = 4{\cdot}222{,}222. \qquad (23.39)$$

Finally, the residual is given by the difference of the total sum of squares and the interactions already found, namely by

$$469{\cdot}319{,}444 - 233{\cdot}736{,}111 - 16{\cdot}152{,}778 - 3{\cdot}444{,}444 - 66{\cdot}097{,}222 - 57{\cdot}888{,}889 - 4{\cdot}222{,}222 = 87{\cdot}777{,}778 \qquad (23.40)$$

with $(6 - 1)(4 - 1)(3 - 1) = 30$ degrees of freedom.

We can now make up the table of variance analysis as follows :—

TABLE 23.11

Analysis of Variance of Data of Table 23.7.

Sum of Squares.		d.f.	Quotient.
Between samples (S) . . .	233·736	5	46·747
,, operators (O) . . .	16·153	3	5·384
,, tests (T)	3·444	2	1·722
Interaction SO	66·097	15	4·406
,, OT	4·222	6	0·704
,, ST	57·889	10	5·789
Residual	87·778	30	2·926
TOTALS	469·319	71	

We proceed to discuss the data in the light of this analysis.

The most striking feature of the table is the size of the quotient between samples. The variance ratio here is $\frac{46{\cdot}747}{2{\cdot}926} = 15{\cdot}976$, with a corresponding value of z equal to 1·38. For $\nu_1 = 5$, $\nu_2 = 30$ the 0·1-per-cent. point is 0·8554, and the ratio is highly significant.

We remark in passing on a point which will be taken up later. The ordinary z-test gives the probabilities that the ratio of two variances chosen *at random* does not exceed a given value. But in this case we have deliberately picked out the largest quotient for one of our estimates. If z had fallen at the 5-per-cent. level we could not have argued that the odds were 19 to 1 against the event. They are very much less, since we have deliberately chosen the largest value for comparison with the residual. However, in the present case our probability is so small that we can confidently assume the significance of z (see **23.27** below).

Our first inference, then, is that the whole sample is not homogeneous. There appear to be variations from sample to sample which are not assignable to differences between tests or operators, and if we wished to standardise our product with greater accuracy we should be led to examine the manufacturing process. This conclusion is, however, subject to a point which we discuss in the next example.

Having rejected the hypothesis of homogeneity we are now faced with the question whether the other quotients in Table 23.11 can be compared so as to assess the relative

variability of the other factors. We must then take a new hypothesis, and we will suppose that the variable may be written

$$x_{jkl} = a_j + \xi_{jkl}, \qquad . \quad . \quad . \quad . \quad . \quad (23.41)$$

where a_j is an unknown quantity expressing the accepted variation between samples. Unless there is something very peculiar about the tests or operators it is reasonable to suppose that the variation between samples can be isolated in this way. We will now suppose that the ξ's, not the x's, are distributed normally with common mean and variance v.

If the values given by (23.41) are substituted in the various constituent items of Table 23.5, it will be found that except for the variation between samples all the other sums of squares assume the same form with ξ written instead of x. This, of course, follows from **23.20** of which our present hypothesis is a particular case. On the hypothesis of (23.41) we are thus enabled to compare the quotients in the table in the usual way. The element of variation between samples has, so to speak, been abstracted from the discussion.

We then turn to the sum of squares between operators in Table 23.11. The variance ratio is $\frac{5{\cdot}384}{2{\cdot}926} = 1{\cdot}84$. For $\nu_1 = 3$, $\nu_2 = 30$ this is not significant. Similarly, for the sum of squares between tests we find a ratio of $\frac{1{\cdot}722}{2{\cdot}926}$, again not significant. Provisionally we conclude that there is no evidence of variation between operators and tests, apart from pure sampling effects.

Now we have to consider the interactions. For that of SO we have the variance ratio $\frac{4{\cdot}406}{2{\cdot}926} = 1{\cdot}51$, which is not significant. We find the same for the interaction ST. For OT we have (taking the larger variance as the numerator)

$$z = \tfrac{1}{2}\log_e \frac{2{\cdot}926}{0{\cdot}703} = 0{\cdot}713, \qquad \nu_1 = 30,\ \nu_2 = 6.$$

This value is just beyond the 5 per cent. point and, judged by itself, might have been regarded as significant ; but taken in conjunction with the others it may, perhaps, be accepted as a permissible sampling fluctuation.

To sum up, therefore, the only evidence of deviation from homogeneity appears in the sample-differences, and we see no reason to reject the hypothesis represented by (23.41). Since all the other items in the analysis, apart from that between samples, are homogeneous, we could condense the table into the form—

Sum of Squares.		d.f.	Quotient.
Between samples	233·736	5	46·747
Remainder	235·583	66	3·569
TOTALS	469·319	71	

The reader may wonder why, in carrying out the tests of significance, we have throughout used the residual quotient as the denominator of the variance ratio, and not, for instance, one of the interactions. There are two reasons. First, the residual has more degrees of freedom, so that it is preferable notwithstanding that the z-test is valid for any number

of degrees of freedom. Second, the residual is not so likely to be affected by interactions which, though not emerging into significance, might nevertheless exist. But once we have established that an interaction is not significant, there is no reason why it should not be amalgamated with the residual, as in the table on page 195.

Example 23.4

There is a point of great importance concerning the inference from analyses of variance, which we will illustrate by an imaginary example based on the data we have just considered. Suppose our analysis of variance were of the following form :—

Sum of Squares.		d.f.	Quotient.
Between samples	125	5	25
Between operators	60	3	20
Interaction SO	150	15	10
Remainder	48	48	1
Totals	383	71	

We will suppose that the sums of squares between tests and the other first-order interactions are not significant, so that they can be amalgamated with the residual to give a remainder with 48 degrees of freedom as shown.

On this evidence the sums of squares between samples and between tests are both significant, as also is the interaction SO. What inference can be drawn about the variability of the product from one sample to another ? We know that the readings differ significantly ; but may not this difference itself be due to the demonstrated variation between operators, or does it really exist ? Is there in fact any variability in the water-content of the product, apart from the sampling effect in homogeneous variation ?

The significance of the SO interaction means that we cannot now regard the effects of operator and sample as independent. We must consider the possibility of entanglement. This is not the only explanation—there may be some other specific cause of variation present which we have not thought of, and on which our present data throw no light. But in this case there is some prior possibility that samples and operators are " entangled " or interacting in the ordinary sense. An operator may be getting better results from his material when it has high water-content than in the reverse case ; or, knowing that the mean content is near 60 per cent. he may unconsciously (or even consciously) bring his determinations nearer to that figure and hence reduce their spread.

In a case of this kind, and indeed in all statistical inquiries, it is important to have a clear idea of the question which is being asked and of the population to which it relates. We have had a number of samples and have tested them by four operators each using three tests. So far as we can see, the tests are equivalent but the operators are not. All the same, we are not very interested in the variation among operators (unless this is an experiment in psychology and not in chemistry). What we want to know is whether the water-content varies in reality, that is to say as the average of a large number of determinations by different operators. Our particular four are themselves samples of a population of operators.

If we confine our attention to the four operators and suppose that each has a specific reaction to particular samples m_{jk}, so that

$$x_{jk} = m_{jk} + \xi_{jk} \qquad . \quad . \quad . \quad . \quad . \quad (23.42)$$

where ξ is a normal random residual with variance v for all j, k, then in the usual way we find

$$E\,\Sigma\,(x_{jk} - x_{j.} - x_{.k} + x_{..})^2 = (p-1)\,(q-1)\,v + \Sigma\,(m_{jk} - m_{j.} - m_{.k} + m_{..})^2 \quad . \quad (23.43)$$

But suppose we consider the matter from a different viewpoint. Regard m_{jk} as itself chosen at random from a normal population of operators with variance v'. Then, taking expectations of this population in addition, we find from (23.43)

$$E\,\Sigma\,(x_{jk} - x_{j.} - x_{.k} + x_{..})^2 = (p-1)\,(q-1)\,(v+v'). \quad . \quad . \quad (23.44)$$

Thus the interaction term provides an unbiassed estimator of the variance $v + v'$ of x_{jk}. By "unbiassed" in this connection we mean that the average over all determinations *and all operators* will give the variance of x_{jk} in the population of all determinations and all operators.

Similarly we shall have, on the same interpretation,

$$\left.\begin{aligned} E\,\Sigma\,(x_{j.} - x_{..})^2 &= (p-1)\,(v+v') \\ E\,\Sigma\,(x_{.k} - x_{..})^2 &= (q-1)\,(v+v') \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (23.45)$$

and hence the ratio of either interaction of zero order *to the first-order interaction* may be tested for homogeneity. Our analysis then becomes—

Sum of Squares.		d.f.	Quotient.
Between samples	125	5	25
Between operators	60	3	20
Residual (SO).	150	15	10
Totals	335	23	

Neither ratio is now significant. For the sum of squares between samples we have a ratio of 2·5, $\nu_1 = 5$, $\nu_2 = 15$, which is below the 5 per cent. point.

Thus we should conclude that, regarding the data as a member of possible samples from all possible operators, there is little or no evidence of real variation from sample to sample. This is quite consistent with the inference we drew at the beginning of the example as to the "significance" of the terms concerned, though at first sight it appears directly contradictory. In the first case we inferred that for these four operators there were significant differences in their determinations for the samples, so that sample-differences are "real" in the sense that they cannot be attributed solely to random variation in homogeneous material. In the second case we enlarge the domain by considering operators as subject to "error" in the sense that one human being differs from another, and find that sample-differences can now be ascribed to variation in the population of operators.

No further emphasis is needed on the care necessary for the proper interpretation of the results of an analysis of variance. The nature of the population which is being considered should be brought explicitly to mind in every case; and the reader should form

the habit of asking himself, whenever a result is found to be "significant": significant of what?

Arithmetic of Variance Analysis

23.26. Before considering further examples we will dispose of a few points arising from the calculation of the constituent sums of squares and the application of the z-test in determining the significance of variance-ratios.

The calculation of sums of squares for an n-way classification can very conveniently be carried out by the use of a punched-card system when the data are numerous, and some remarkable computing feats have been performed by this technique. For ordinary laboratory work with a machine, the process of Example 23.3 is possibly the best, though some modifications may be made to suit individual taste.

The main work lies in computing the total sum of squares. This is done by finding the sum of squares of observations from the original data (with a convenient working mean) and the sum of observations obtained at the same time. The formula

$$\begin{aligned}\Sigma\,(x_{jkl} - x_{\dots})^2 &= \Sigma\, x_{jkl}^2 - N x_{\dots}^2 \\ &= \Sigma\, x_{jkl}^2 - x_{\dots}\, \Sigma\, x_{jkl} \qquad (23.46)\end{aligned}$$

then gives the total sum required. The quantity $Nx_{\dots}^2$ is constantly needed and should be recorded. It is useful to preserve a few more decimal places than will ultimately be used in the final presentation of the analysis.

The original data are then condensed into n $(n-1)$-way tables by summing over each class in turn. In Example 23.3 this was done so as to give three tables: Operators–Samples, Tests–Samples and Operators–Tests. The main body of these tables gives means of the type $x_{jk.}$ multiplied by a constant factor. A further condensation will give $\binom{n}{2}$ sets of means of type $x_{j..}$; and so on, as far as is required.

From the condensed tables we can then determine the sums of squares of means of various orders, and hence the interactions. The main pitfall lies in the way of the application of the correct multipliers and divisors—it has to be borne in mind that the summation takes place over all values of the sample.

Suppose, for example, we have a four-way classification into classes with p, q, r and s numbers of members. The first condensation gives us four tables of which a typical one is $p \times q \times r$, based on the sum of s members. The next condensation gives us six two-way tables typified by $p \times q$, based on the sum of rs members. The third gives us four one-way tables such as p, based on qrs members. Consider the variance between p-classes :—

$$\Sigma\,(x_{j\dots} - x_{\dots\dots})^2 = \Sigma\, x_{j\dots}^2 - N x_{\dots\dots}^2 \qquad (23.47)$$

In the condensed one-way table of p classes each term is to be counted qrs times, and thus, if S is the sum of squares in this table as it stands,

$$S = \sum_{j=1}^{p} (qrs\, x_{j\dots})^2.$$

Thus, summing over all members, we find

$$\begin{aligned}\Sigma\, x_{j\dots}^2 &= qrs \frac{S}{(qrs)^2} \\ &= \frac{S}{qrs}, \qquad (23.48)\end{aligned}$$

whence (23.47) gives the zero-order interaction for p-classes. Similarly for q, r and s.

For the first-order interaction we have

$$\Sigma (x_{jk..} - x_{j...} - x_{.k..} + x_{....})^2 = \Sigma (x_{jk..} - x_{....})^2 - \Sigma (x_{j...} - x_{....})^2 - \Sigma (x_{.k..} - x_{....})^2 . \qquad (23.49)$$

The last two terms on the right have already been found. We require

$$\Sigma (x_{jk..} - x_{....})^2 = \Sigma x_{jk..}^2 - N x_{....}^2 \qquad (23.50)$$

If S' is the sum of squares of elements in the body of the two-way table found by adding r- and s-items, we find

$$\Sigma x_{jk..}^2 = \frac{S'}{rs}, \qquad (23.51)$$

and so on. The general process will now be clear.

Unfortunately there is no convenient independent check on the calculations. The various condensed tables are self-checking since their totals are the sum of all observations, but the sums of squares do not check with anything. It is, of course, possible to evaluate each individual term in the residual and to check by summing squares, but this is too laborious for use except in the simplest cases.

Use of the z-test for Several Variance-ratios

23.27. In the complete analysis of n classes there are $2^n - 1$ elements, and the number of variance ratios arising for test may be considerable. The z-test gives the probability that a particular value chosen at random will be exceeded. If therefore we pick out the largest ratios for test, the chance that one of them is " significant " in the sense of exceeding the $100P$-per-cent. point is a good deal greater than P, and we run into the danger of attributing significance to what may be a pure sampling effect.

Suppose we make r different *and independent* tests of r values of z. The chance that each does not exceed a fixed value (depending on the number of degrees of freedom) is $1 - P$, where P is some assigned level of significance. Hence the chance that none of them exceeds its appropriate value is

$$(1 - P)^r = 1 - rP, \text{ approximately}, \qquad (23.52)$$

provided that P and rP are small. For instance, if $P = 0{\cdot}01$ and $r = 7$ the probability that no z exceeds its appropriate significance value is $0{\cdot}93$, and thus there is a probability of $0{\cdot}07$ that at least one of them will do so.

In practice the problem of numerous comparisons is more complicated because they are not independent. In such circumstances our judgment of significance has to incorporate an element of the intuitive. However, if all the comparisons are based on the common residual quotient it is possible to find the probabilities that the largest of r values exceeds assigned values. The resulting expressions are complicated, even when all the sums of squares have the same degrees of freedom, but reference may be made to Hartley (1938) for approximations and to Cochran (1941) and Finney (1941*a*) for exact expressions. The conclusion reached by Finney is that if the degrees of freedom in the residual are sufficiently numerous the ratios may be treated as completely independent.

23.28. There is a particular case of the n-way classification which is worth special mention, namely, that for which each classification is a simple dichotomy, so that there are 2^n subgroups. This case arises frequently when so-called " factorial " experiments are being conducted to determine the effect of a treatment which is either applied or with-

held. The analysis of variance remains the same in principle, but of course the arithmetic becomes a good deal simpler.

Example 23.5 (F. Yates, *Supp. J.R.S.S.*, 1935, **2**, 181)

An area of ground was sown with peas and divided into 24 plots in the manner shown in Table 23.12. The plots received, or did not receive, dressings of nitrogen (N), phosphate (P) and potash (K) in the manner shown, the yields in pounds being given in the table.

TABLE 23.12

Yields of Peas and Manurial Treatments on 24 Plots

PK 49·5	46·8	N 62·0	K 45·5
NP 62·8	NK 57·0	NPK 48·8	P 44·2
N 59·8	K 55·5	NP 52·0	NK 49·8
NPK 58·5	P 56·0	51·5	PK 48·8
P 62·8	N 69·5	NK 57·2	PK 53·2
NPK 55·8	K 55·0	NP 59·0	56·0

There is some purpose here in the alternation of treatments, but that need not concern us for the present. We have 24 observations in four classes, viz. blocks (3), nitrogen (2), phosphate (2) and potash (2), giving $3 \times 2 \times 2 \times 2 = 24$ records.

Condensing the table by adding blocks we get the following :—

No treatment	N	P	K	NP	NK	PK	NPK	TOTAL
154·3	191·3	163·0	156·0	173·8	164·0	151·5	163·1	1317·0

Condensing according to the three treatments we have—

	N	not-N	TOTALS
P	336·9	314·5	651·4
not-P	355·3	310·3	665·6
TOTALS	692·2	624·8	1317·0

	K	not-K	TOTALS
P	314·6	336·8	651·4
not-P	320·0	345·6	665·6
TOTALS	634·6	682·4	1317·0

	N	not-N	TOTALS
K	327·1	307·5	634·6
not-K	365·1	317·3	682·4
TOTALS	692·2	624·8	1317·0

We omit the remaining calculations. The analysis in its final form is given in ble 23.13.

TABLE 23.13

Analysis of Variance of the Data of Table 23.12

Sums of Squares.		d.f.	Quotient.
Between blocks (B)	177·803	2	88·90
,, N	189·282	1	189·28
,, P	8·402	1	8·40
,, K	95·202	1	95·20
Interaction BN	94·255	2	47·13
,, BP	2·260	2	1·13
,, BK	23·685	2	11·84
,, NP	21·281	1	21·28
,, NK	33·134	1	33·13
,, PK	0·481	1	0·48
,, BNP	25·302	2	12·65
,, BNK	36·004	2	18·00
,, BPK	3·782	2	1·89
,, NPK	37·003	1	37·00
Residual ($BNPK$)	128·489	2	64·24
TOTALS	876·365	23	

We have carried out the analysis in full so as to illustrate the arithmetical process a four way classification, but we may note at once that it is unduly elaborate. There only 24 observations in the data and we cannot expect them to provide all the answers the questions which we could frame as to the significance of the various constituent ms in the analysis. This is borne out by the z-test. The residual variance is 64·24 th two degrees of freedom. For $\nu_1 = 1$, $\nu_2 = 2$ the variance ratio at the 1-per-cent. int is 98·49 and that for $\nu_1 = 2$, $\nu_2 = 2$ at the same point is 99·00. Only values greater an about 100 times 64·24 or less than 1/100th of that value would thus be significant. ly the interaction PK falls outside this range, and even this, among so many, can hardly regarded as significant.

The inquiry is not, however, completely frustrated. Since the second-order inter- tions are not significant, we amalgamate them with the residual to give a remainder m of squares of 230·580 with nine d.f. and a quotient of 25·62. It will now be found

that among the first-order interactions only two are significant, PK and BP being too small. Had they been too large we might have attributed some genuine significance to this result, but it is not very plausible to suppose that there is a " real " interaction between blocks and phosphate, or that phosphate and potash inhibit each other's action. The differences from expectation are more probably due to individual soil variation from plot to plot.

If we accept the first-order interactions as not significant, we may amalgamate them with the remainder to give the following :—

Sum of Squares.		d.f.	Quotient.
Blocks	177·803	2	88·90
N	189·282	1	189·28
P	8·402	1	8·40
K	95·202	1	95·20
Remainder	405·676	18	22·54
TOTALS	876·365	23	

Here the P-quotient is not significant, but the variance ratio for blocks, 3·99, is near the 5-per-cent. point. The N-quotient will be found to be significant at the 1-per-cent. point, the K-quotient near to the 5-per-cent. point. Our conclusion is that there is strong indication that nitrogen influenced the yield, some indication that potash did so, and little indication that phosphates did so ; and that there is ground for suspecting heterogeneity in the soil partly because of the difference between blocks and partly from some of the first-order interactions.

In this case, of course, we knew already more or less what was to be expected of these data and are the readier to accept the conclusions on that account. Had we known nothing of the effect of fertilisers on leguminous crops our conclusions on such slender evidence must have been very tentative indeed, particularly if we wished to extend them to peas grown on other soils under different climatic conditions with different amounts of fertiliser.

Example 23.6 (C. E. Gould and W. M. Hampton, *Supp. J.R.S.S.*, 1936, **3**, 137)

In the manufacture of optical glass there appear small bubbles known as " seed ", which constitute a defect. The glass is made in " pots " which take about a year to prepare, and are run continuously over long periods when once started. There are two pots to a furnace and materials are introduced into a pot from time to time which, after fusion, provide a " run " of glass. Each run provides several days' work, one day's work being known as a " journey ". At each journey quantities of glass are drawn from the pot and blown into " cylinders ", there being about 18 or 20 to the journey. For the purposes of the experiment three cylinders were chosen, the third, tenth and sixteenth, and pieces of regular size cut from them for examination as to frequency of seed. The first five journeys of each of five runs were sampled.

We have here a four-way classification 2 (pots) $\times$ 5 (runs per pot) $\times$ 5 (journeys per run per pot) $\times$ 3 (cylinders per journey per run per pot). The actual dates of the runs were February 16th, May 23rd, June 12th, September 1st and December 6th, so that the manufacturing period covered about ten months. We shall assume that the glass was

of the same type throughout, although in actual fact it was different in one or two cases —but not sufficiently different to affect the analysis.

The topic of main interest here is whether the frequency of seed varies significantly according to the four factors concerned. If so, the alteration of manufacturing conditions may improve the wastage due to seed; but if not—and the variation is the kind of thing which can be accounted for as chance fluctuation in sampling from a homogeneous population—there is little hope of improvement except perhaps by a radical alteration in the process affecting all pots, runs and journeys alike.

TABLE 23.14

Frequency of "Seed" in Samples of Glass

		Pot 1.			Pot 2.	
	Cyl. 1.	Cyl. 2.	Cyl. 3.	Cyl. 1.	Cyl. 2.	Cyl. 3.
Run 1 J 1 . . .	47	56	100	52	61	88
2 . . .	55	89	93	49	62	97
3 . . .	35	57	56	34	60	72
4 . . .	78	67	113	47	93	118
5 . . .	33	40	128	16	29	130
Run 2 J 1 . . .	52	66	36	65	80	40
2 . . .	21	61	49	122	97	79
3 . . .	31	39	25	45	54	72
4 . . .	43	72	52	109	120	80
5 . . .	37	51	67	67	85	63
Run 3 J 1 . . .	50	61	60	75	139	130
2 . . .	33	27	49	46	58	63
3 . . .	24	39	24	15	33	39
4 . . .	18	18	43	22	16	19
5 . . .	28	42	28	27	19	22
Run 4 J 1 . . .	24	34	43	46	66	24
2 . . .	24	49	42	40	117	105
3 . . .	21	21	51	30	28	34
4 . . .	21	69	48	36	64	53
5 . . .	76	48	42	39	60	78
Run 5 J 1 . . .	31	54	40	19	93	36
2 . . .	34	24	46	16	12	2
3 . . .	120	122	120	33	58	107
4 . . .	109	119	120	25	63	90
5 . . .	69	49	60	34	43	30

Before plunging into the analysis of variance it is as well to look over the data to see whether they themselves suggest any lines of inquiry. We observe considerable variability from journey to journey within the same run, $J3$ and $J4$ of run 5 being conspicuous in pot 1; and in run 1 the numbers of seed appear to increase from cylinder 1 to cylinder 3 in a rather exceptional way. The runs themselves seem to differ materially. Prior con-

siderations also suggested an examination of the way in which frequency of seed varied between pots, since they were chosen so as to differ substantially in constitution.

A complete analysis of variance of the data is as follows :—

TABLE 23.15

Analysis of Variance of the Data of Table 23.14.

Sums of Squares.		d.f.	Quotient.
Between pots (P)	898	1	898
„ runs (R)	14,059	4	3,515
„ journeys (J)	4,355	4	1,089
„ cylinders (C)	10,631	2	5,315
Interaction PR	16,133	4	4,033
„ PJ	4,081	4	1,020
„ PC	587	2	293
„ RJ	45,934	16	2,871
„ RC	11,626	8	1,453
„ JC	2,540	8	317
„ PRJ	9,711	16	607
„ RJC	12,472	32	390
„ JCP	1,656	8	207
„ CPR	1,862	8	233
Residual ($PRJC$)	8,110	32	253
Totals	144,655	149	

The second-order interactions will be found non-significant, so we amalgamate with the residual, giving a sum of squares 33,811, d.f. 96, quotient 352.

It then appears that of the first-order interactions PR, RJ and RC are significant and PJ may be so. There is beginning to appear evidence of heterogeneity, and that of a rather complicated kind. It seems that pots are interacting with runs, runs with journeys and runs with cylinders.

Taking 352 as the quotient, we find that except for P the zero-order interactions are significant. The five R-means are 68·50, 62·67, 42·23, 47·77 and 59·27, so that the variation of runs is not a simple rise or fall, which could have been explained as a time-effect. The five J-means are 58·93, 55·37, 49·97, 64·83 and 51·33, again not a regular effect. The C-means are 44·46, 59·68 and 64·12, which are significantly different. Inspection of the table suggests that the first run is the source of the trouble.

With data as heterogeneous as these it is rather difficult to set up a plausible hypothesis to test. The interactions of first order suggest that no simple additive effects of the four factors will explain observation, and if these terms are used as denominators in tests of variance ratios the variation between classes appears on the whole non-significant on the usual hypotheses. The analysis, then, suggests several subjects for inquiry as concerns the homogeneity of the data, but does not suggest any simple explanation of the observed figures. The reader may care to refer to the original paper for a more complete discussion of the subject.

23.29. Perhaps we may pause at this point to review progress. We have seen that for an n-way classification of the special type wherein each subclass contains a single member, the sum of squares of all observations about their mean can be exhibited as the sum of a number of such sums. On the hypothesis of normality and homogeneity each constituent sum of squares, on division by its appropriate number of degrees of freedom, gives an estimator of the parent variance, and each is distributed as χ^2 independently of the others. The hypothesis of homogeneity can then be tested in Fisher's z-distribution, subject to the adoption of a conservative attitude where many tests are made on the same data. If the hypothesis is rejected we may replace it by a simple form in which the effects of the different classes are additive, provided that the interactions are not significant. The particular ratio chosen for a test depends on the hypothesis concerned, and it is important to have a clear idea of the exact question to which an answer is sought.

23.30. In the next chapter we shall consider the case when the numbers in different subclasses are not equal, discuss the additive hypothesis in more detail, examine the relationship of variance- and regression-analysis, and extend our results to the analysis of covariance. We conclude this chapter by an examination of the important question: what can be done with the analysis of variance when the variation is not normal?

Non-normal Data

23.31. The analysis of a sum of squares into its constituent sums can, of course, be undertaken in all circumstances, but the various quotients may not continue to provide unbiassed estimators of the parent variance if the population is not-normal. What is equally serious, the constituent sums of squares may not be distributed independently. Thus, when parent normality cannot be assumed, the quotients in the analysis table are no longer equal within sampling limits and their ratio is distributed in unknown form; and even if the form were known it would probably depend on parent parameters and hence fail to provide an exact test of significance.

The problem has been considered in four ways:—

(*a*) Sampling experiments have been undertaken to see how far moderate deviation from normality affects the z-distribution;

(*b*) Attempts have been made to find transformations of the variate to throw the parent distributions into forms with equal variances, at least approximately, before the analysis is applied;

(*c*) By introducing a randomising process into the data before they are collected, attempts have been made to preserve the z-distribution as a close approximation —this amounts to a change in the nature of the inference, as we shall see below;

(*d*) Tests have been found which can be applied to ranked data irrespective of the parent form—this approach is a particular case of (*c*), but seems to merit special mention.

We proceed to consider these four possibilities.

23.32. The arithmetic entailed by a single analysis of variance, even in simple cases, implies that an extensive sampling inquiry into the distribution of z in non-normal populations would be a very formidable undertaking. E. S. Pearson (1931*b*) has studied in some detail the case of a one-way classification with unequal numbers, when the distribution

of z becomes equivalent to that of the correlation ratio η^2. Six populations were chosen, characterised by the following values :—

$$\begin{aligned}
&\beta_1 = 0, &&\beta_2 = 2{\cdot}50 \text{ (symmetrical platykurtic)};\\
&\beta_1 = 0, &&\beta_2 = 4{\cdot}1 \text{ (symmetrical leptokurtic)};\\
&\beta_1 = 0, &&\beta_2 = 7{\cdot}05 \text{ (symmetrical leptokurtic)};\\
&\beta_1 = 0{\cdot}2, &&\beta_2 = 3{\cdot}3 \text{ (skew, Type III)};\\
&\beta_1 = 0{\cdot}49, &&\beta_2 = 3{\cdot}72 \text{ (skew, Type III)};\\
&\beta_1 = 0{\cdot}99, &&\beta_2 = 3{\cdot}83 \text{ (very skew, Type I, with abrupt start)}.
\end{aligned}$$

The results suggested that for this range of β_1 and β_2 the distribution of z is adequately represented by Fisher's distribution, and that therefore the homogeneity test may be applied. The case when the variation changed from group to group was not considered. It was also concluded that " it seems probable that the more elaborate forms of analysis of variance are also of fairly wide application ".

Some work by Eden and Yates (1933) is often referred to as experimental confirmation of the same kind, but in fact it was carried out with rather a different object, that of confirming the z-test for data under randomisation (see below, **23.36**).

Variate Transformations

23.33. Suppose ξ is a new variate $\xi(x)$. Then approximately we shall have

$$\operatorname{var} \xi = \left(\frac{d\xi}{dx}\right)^2 \operatorname{var} x. \qquad (23.53)$$

If now the parent variance of the x-distribution is related in some known manner to the mean, say $f(m) = v$, we have

$$\operatorname{var} \xi = \left(\frac{d\xi}{dx}\right)^2 f(m).$$

As a further approximation, if x varies about m by small quantities we have

$$\operatorname{var} \xi = \left(\frac{d\xi}{dx}\right)^2 f(x). \qquad (23.54)$$

Now we wish ξ to have a constant variance, say λ, and if this is so,

$$\frac{d\xi}{dx} = \sqrt{\frac{\lambda}{f(x)}},$$

or

$$\xi = \int \sqrt{\frac{\lambda}{f(x)}}\, dx. \qquad (23.55)$$

Although this expression is arrived at by approximation we are entitled to hope that the variate ξ will have almost constant variance, and at any rate a more stable variance than x.

For instance, if the original variation is thought to be of the Poisson type we have $f(x) = x$, and from (23.55) are led to consider the transformation

$$\begin{aligned}\xi &= \int \frac{\sqrt{\lambda}}{\sqrt{x}}\, dx \\ &= \sqrt{x}, \qquad (23.56)\end{aligned}$$

if we choose λ to be $\frac{1}{4}$. Similarly, if the variation is of the binomial type with variance $p(1-p)$ we have

$$\xi = \int \frac{\sqrt{\lambda}}{\sqrt{\{p\,(1-p)\}}}\,dp$$

$$= \sin^{-1}\sqrt{x}, \qquad \qquad (23.57)$$

on suitable choice of λ.

23.34. These transformations are designed to "stabilise" the variance. They do not necessarily bring the variate closer to normality, though in some cases they will do so —we have, for instance, seen that $\sqrt{\chi^2}$ tends to normality quicker than χ^2 (**12.7**). The following values (Bartlett 1936*d*) illustrate the way in which the square-root transformation stabilises the variance of a Poisson distribution :—

Mean m.	Variance of Poisson Variate $\sqrt{x}$.	Variance of Poisson Variate $\sqrt{(x+\frac{1}{2})}$.
0·0	0·000	0·000
0·5	0·310	0·102
1·0	0·402	0·160
2·0	0·390	0·214
3·0	0·340	0·232
4·0	0·306	0·240
6·0	0·276	0·245
9·0	0·263	0·247
12·0	0·259	0·248
15·0	0·256	0·248

The term $\frac{1}{2}$ in the third column was added by Bartlett on the analogy of a continuity correction. For $m > 3$ the variance is evidently quite stable.

23.35. If now, having stabilised the variance, we carry out an analysis in the ordinary way, our residual sums of squares divided by the appropriate degrees of freedom will continue to be unbiassed estimates of the common variance v, even if there are differences between the means of the classes. Instead of assuming as part of the hypothesis that the different classes are distributed with the same variance, we have transformed the variate so that this shall be so, at least to a close approximation. Relying further on the result that the transformed variates approximate to normality, or that if they do not the difference will not seriously vitiate the z-test, we may apply that test to the transformed data in the usual way.

Example 23.7 (Bartlett, 1936*d*)

Table 23.16 shows the number of wheat seeds out of 50 which failed to germinate in four repetitions of an experiment with different treatments.

TABLE 23.16

Germination of Wheat Seeds

Number of Experiment.	Number of Treatment.							Totals.
	1	2	3	4	5	6	7	
1	10	11	8	9	7	6	9	60
2	8	10	3	7	9	3	11	51
3	5	11	2	8	10	7	11	54
4	1	6	4	13	7	10	10	51
Totals	24	38	17	37	33	26	41	216

In point of fact, treatment 7 was a repetition of treatment 6, the others being different. The point of interest is whether the treatments exert any effect on germination. We shall not inquire into any differences between experiments (which appear to be negligible from the row totals) and shall accordingly consider this as a one-way classification into seven classes, four numbers to the class.

The presumption is that in any given class the variation is of the binomial type. We might apply the $\sin^{-1}\sqrt{x}$ transformation, but will adopt instead an *ad hoc* square-root transformation obtained as follows :—

We have

$$v = np(1-p).$$

Suppose now that $p = p_0 + \delta$ where δ is small. Then

$$\begin{aligned} v &= n(p_0 + \delta - p_0^2 - 2p_0\delta) \\ &= n\{(1-2p_0)(p-p_0) + p_0 - p_0^2\} \\ &= np(1-2p_0) + np_0^2. \end{aligned}$$

If we now put

$$\xi = \sqrt{(x + k + \tfrac{1}{2})}$$

where $k = \dfrac{np_0^2}{1-2p_0}$ and x is the observed frequency, then ξ will tend to have constant variance.

In our example the total frequency is 216 out of 1400 seeds, so that we may take as an estimate of p_0 the ratio $216/1400 = 0{\cdot}15$. The transformed variate then becomes

$$\xi = \sqrt{\left\{np + \tfrac{1}{2} + \frac{50\,(\cdot 0225)}{0{\cdot}70}\right\}}$$

$$= \sqrt{(np+2)}, \text{ approximately.}$$

On this basis the transformed variate-values are—

TABLE 23.17

Transformed Variates of Table 23.16

Number of Experiment.	Number of Treatment. 1	2	3	4	5	6	7	TOTALS.
1	3·464	3·606	3·162	3·317	3·000	2·828	3·317	22·694
2	3·162	3·464	2·236	3·000	3·317	2·236	3·606	21·021
3	2·646	3·606	2·000	3·162	3·464	3·000	3·606	21·484
4	1·732	2·828	2·449	3·873	3·000	3·464	3·464	20·810
TOTALS	11·004	13·504	9·847	13·352	12·781	11·528	13·993	86·009

The analysis of variance is—

Sums of Squares.		d.f.	Quotient.
Between treatments	3·486	6	0·581
Residual	4·316	21	0·206
TOTALS	7·802	27	

The sum of squares is particularly easy to obtain, being the sum of the original variates plus twice the number of variate-values.

The variance ratio, 2·8, is barely significant, being just beyond the 5-per-cent. point. There is little evidence that treatments are exerting any effect on germination, since a comparison of treatments 6 and 7 (which are the same) indicates that such "significance" as exists may be due to heterogeneity in the seed.

Randomisation

23.36. Consider a two-way classification of pq members, the observed value of the jth A-member of the kth B-class being x_{jk}. Following the line already considered in **21.48**, we will consider the z-distribution in the population of values obtained by permuting the members in any A-class in all possible ways. There will thus be $(q\,!)^p$ possible values of z, all based on the observed values. We have already considered a case of this kind in dealing with the problem of m rankings (**16.29**) and we shall follow the same procedure in solving the more general problem.

Let the values be arrayed as

$$\left.\begin{array}{ccccc} x_{11} & x_{12} & \cdot & \cdot & x_{1q} \\ x_{21} & x_{22} & \cdot & \cdot & x_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{p1} & x_{p2} & \cdot & \cdot & x_{pq} \end{array}\right\} \qquad (23.58)$$

If S_R is the sum of squares between rows, S_C that between columns and S the total, we know that in the ordinary case considered earlier in the chapter, S_C is distributed as $v\chi^2$ with $q - 1$ d.f., and $S - S_R - S_C$ as $v\chi^2$ with $(p - 1)(q - 1)$ d.f. It follows that

$$\frac{S_C}{S - S_R} = W, \text{ say,} \qquad (23.59)$$

is distributed in the Type I form

$$dF \propto W^{\frac{1}{2}(q-1)-1} (1 - W)^{\frac{1}{2}(p-1)(q-1)-1}\, dW. \qquad (23.60)$$

It is easier to work with W than with z, but there is of course no difficulty in passing from one to the other.

We proceed to find the first four moments of W in the population of $(q\,!)^p$ values obtained by permuting the rows of (23.58) in all possible ways.

23.37. If in (23.58) we increase the members of any row by a constant a, it is easily seen that S_C and $S - S_R$ remain unaffected, and hence so does W. Thus we may take the mean of each row to be zero and then $S_R = 0$. With this origin we have

$$W = \frac{S_C}{S} = \frac{\sum\limits_i \left(\sum\limits_j x_{ij}\right)^2}{\sum\limits_{i,j} x_{ij}^2} \qquad (23.61)$$

If now

$$R_{ik} = \sum_{j=1}^{q} (x_{ij}\, x_{kj}) \qquad (23.62)$$

and the k-statistics of the q values x_{ij}, $j = 1 \ldots q$, are written k_{i1}, k_{i2}, etc., and

$$U = \sum_{i,k} R_{ik}, \qquad (23.63)$$

we find

$$W = \frac{1}{p} + \frac{2U}{p\,(q - 1) \sum\limits_i k_{i2}} \qquad (23.64)$$

$$E(R_{ik}) = 0 \qquad (23.65)$$

$$E(R_{ik}^2) = (q - 1)\, k_{i2}\, k_{k2} \qquad (23.66)$$

$$E(R_{ik}^3) = \frac{(q - 1)(q - 2)}{q}\, k_{i3}\, k_{k3} \qquad (23.67)$$

$$E(R_{ik}^4) = \frac{3\,(q - 1)^3}{q + 1}\, k_{i2}^2\, k_{k2}^2 + \frac{(q - 1)(q - 2)(q - 3)}{q\,(q + 1)}\, k_{i4}\, k_{k4}. \qquad (23.68)$$

Then, for the moments of U,

$$E(U) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.69)$$

$$E(U^2) = (q-1)\sum_{i,k}{}' k_{i2}\, k_{k2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.70)$$

$$E(U^3) = 6(q-1)\sum_{i,k,l}{}' k_{i2}\, k_{k2}\, k_{l2} + \frac{(q-1)(q-2)}{q}\sum_{i,k}{}' k_{i3}\, k_{k3} \quad . \quad . \quad . \quad (23.71)$$

$$\begin{aligned} E(U^4) = {} & \frac{3(q-1)^3}{q+1}\sum_{i,k}{}' k_{i2}^2\, k_{k2}^2 + \frac{(q-1)(q-2)(q-3)}{q(q+1)}\sum_{i,k}{}' k_{i4}\, k_{k4} \\ & + 3(q-1)^2\{(\Sigma' k_{i2}\, k_{k2})^2 - \Sigma' k_{i2}^2\, k_{k2}^2\} \\ & + \frac{12(q-1)(q-2)}{q}\Sigma' k_{i3}\, k_{k3}\, k_{l2} + 72(q-1)\Sigma' k_{i2}\, k_{k2}\, k_{l2}\, k_{m2} \quad . \quad (23.72) \end{aligned}$$

where Σ' denotes summation over values for which the subscripts are unequal and permutations are not allowed.

Finally, for the moments of W we have

$$E(W) = \frac{1}{p} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.73)$$

$$E(W-\bar{W})^2 = \frac{4}{p^2(q-1)}\frac{\Sigma' k_{i2}\, k_{k2}}{(\Sigma k_{i2})^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.74)$$

$$E(W-\bar{W})^3 = \frac{48}{p^3(q-1)^2}\frac{\Sigma' k_{i2}\, k_{k2}\, k_{l2}}{(\Sigma k_{i2})^3} + \frac{8(q-2)}{p^3 q(q-1)^2}\frac{\Sigma' k_{i3}\, k_{k3}}{(\Sigma k_{i2})^3} \quad . \quad . \quad . \quad (23.75)$$

$$\begin{aligned} E(W-\bar{W})^4 = {} & \frac{48}{p^4(q-1)^2}\frac{(\Sigma' k_{i2}\, k_{k2})^2}{(\Sigma k_{i2})^4} - \frac{96}{p^4(q-1)^2(q+1)}\frac{\Sigma' k_{i2}^2\, k_{k2}^2}{(\Sigma k_{i2})^4} \\ & + \frac{1152}{p^4(q-1)^3}\frac{\Sigma' k_{i2}\, k_{k2}\, k_{l2}\, k_{m2}}{(\Sigma k_{i2})^4} \\ & + \frac{16(q-2)(q-3)}{p^4(q+1)(q)(q-1)^3}\frac{\Sigma' k_{i4}\, k_{k4}}{(\Sigma k_{i2})^4} + \frac{192(q-2)}{p^4(q-1)^3 q}\frac{\Sigma' k_{i3}\, k_{k3}\, k_{l2}}{(\Sigma k_{i2})^4} \quad . \quad (23.76) \end{aligned}$$

These formulae can be derived in the manner of **16.33**, but reference may be made to Pitman (1938) for further details.

23.38. We now consider how far the first four moments of W, as found above, agree with the first four moments of the distribution (23.60). The mean and variance of the latter are

$$\frac{1}{p} \quad \text{and} \quad \frac{2(p-1)}{p^2(pq-p+2)}. \quad . \quad . \quad . \quad . \quad . \quad (23.77)$$

The means agree exactly. For the variances to agree we must have, from (23.74) and (23.77),

$$\frac{4}{p^2(q-1)}\frac{\Sigma' k_{i2}\, k_{k2}}{(\Sigma k_{i2})^2} = \frac{2(p-1)}{p^2(pq-p+2)}. \quad . \quad . \quad . \quad (23.78)$$

Writing

$$K = \frac{2\,\Sigma' k_{i2}\, k_{k2}}{(\Sigma k_{i2})^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23.79)$$

we find that (23.78) is equivalent to

$$K = \frac{(p-1)(q-1)}{pq - p + 2}. \qquad (23.80)$$

The ratio K may have any value from 0 to $\frac{p-1}{p}$, the lower limit being approached when one of the second k-statistics k_{i2} is much larger than the others, the upper limit when they are all equal. Hence all that can be said about the variance of W is that it is not greater than $\frac{2(p-1)}{p^3(q-1)}$ and that it takes this value when the variance of each p-class is the same.

Turning to the third and fourth moments, we note that in many cases where the variation is not too skew the quantities k_{i3} and k_{i4} will be negligible. A number of terms in (23.75) and (23.76) may thus be neglected, but even those that remain are fairly complicated, and it is difficult to say how far the distribution of W will approach the Type I distribution (23.60). In practice the values may be worked out and compared. If there is reasonable agreement, the z-distribution of the variance ratio will hold in the particular population which we are considering.

23.39. A better approach is to find the Type I distribution which has the same first two moments as W and to modify the z-test where necessary. It may be shown that when K is not too small the third and fourth moments of W and the fitted Type I distribution are in fairly good agreement, so that we may expect a good fit.

The Type I distribution with mean $\frac{1}{p}$ and variance $\frac{2K}{p^2(q-1)}$ has the mean and variance of W by definition. Its third moment is easily seen to be

$$\frac{8K^2}{p^3(q-1)} \frac{p-2}{p-1+\frac{2K}{q-1}}. \qquad (23.81)$$

We have to see how far this differs from the actual third moment of W given by (23.75).

Now

$$\begin{aligned} 3\,\Sigma' k_{i2} k_{k2} k_{l2} &= \Sigma k_{i2}\, \Sigma' k_{k2} k_{l2} - \Sigma' k_{i2}^2 k_{k2} \\ &= \Sigma k_{i2}\, \Sigma' k_{k2} k_{l2} - (\Sigma k_{i2}\, \Sigma k_{i2}^2 - \Sigma k_{i2}^3) \\ &= \Sigma k_{i2}\, (3\Sigma' k_{i2} k_{k2} - \Sigma k_{i2}^2) + \Sigma k_{i2}^3, \end{aligned}$$

and hence

$$\frac{6\,\Sigma' k_{i2} k_{k2} k_{l2}}{(\Sigma k_{i2})^3} = 3K - 2 + 2\frac{\Sigma k_{i2}^3}{(\Sigma k_{i2})^3}. \qquad (23.82)$$

Since all the k's here concerned are positive,

$$\Sigma k_{i2}\, \Sigma k_{i2}^3 \geqslant (\Sigma k_{i2}^2)^2$$

and hence

$$\frac{\Sigma k_{i2}^3}{(\Sigma k_{i2})^3} \geqslant \left\{ \frac{\Sigma k_{i2}^2}{(\Sigma k_{i2})^2} \right\}^2 = (1-K)^2. \qquad (23.83)$$

Hence, from (23.82) and (23.83),

$$6\frac{\Sigma' k_{i2} k_{k2} k_{l2}}{(\Sigma k_{i2})^3} \geqslant 3K - 2 + 2(1-K)^2 = K^2\left(1 - \frac{1-K}{K}\right). \qquad (23.84)$$

Similarly, since

$$\frac{\Sigma k_{i2}^3}{(\Sigma k_{i2})^3} \leqslant \left\{\frac{\Sigma k_{i2}^2}{(\Sigma k_{i2})^2}\right\}^{\frac{3}{2}} = (1-K)^{\frac{3}{2}} < (1-K)\,(1-\tfrac{1}{2}K-\tfrac{1}{8}K^2)$$

it appears that

$$6\,\frac{\Sigma' k_{i2}\,k_{k2}\,k_{l2}}{(\Sigma k_{i2})^3} < K^2\,\frac{3+K}{4}. \qquad (23.85)$$

On comparing (23.75) and (23.81), and assuming that the second term in the former may be neglected, we see that they differ by the factor whose limits we have found in (23.84) and (23.85), namely

$$1-\frac{1-K}{K} \quad \text{and} \quad \frac{3+K}{4}.$$

If K is not too small the limits are not very different from unity, and the third moments are accordingly in fairly good agreement.

In the same way but with rather more complicated algebra it may be shown that the fourth moments are in fair agreement.

When all the rows are rankings, the case reduces to that considered in **16.33** *et seq.*, and we have already seen that the distribution of W is closely approximated by the Type I distribution in that case.

23.40. Suppose, now, that we have p classes of objects, one of each class belonging to a second series of classes, q in number. As our hypothesis we will suppose that membership of the q-classes is independent of the variate-values, so that we may suppose it to be a matter of chance how the values in any p-class are distributed among the q-classes. On this hypothesis the variance ratio will follow the z-form approximately (subject to the conditions we have discussed above) in the population consisting of the $(q\,!)^p$ permutations of *observed values*; and this will be so whether the parent is normal or not.

By shaping the inference in this way, and making it conditional, we are thus able to apply the z-test even in cases of non-normality. The test of homogeneity still applies, but of course the inference is rather different from the usual type. This point has not, perhaps, been adequately emphasised in the past and there still seems to be confusion on the subject.

Randomised Blocks

23.41. The principle of testing in a conditional population has received its chief applications in a certain type of agricultural experiment (and analogous cases in other fields), known as a randomised block experiment. We are given p blocks of land and wish to test the existence of differential effects among q treatments, e.g. manurial treatments, of a crop to be grown on it. We divide each block into q plots and grow the crops on each of the pq plots. In any one block we apply a different treatment to each of the q plots; and we allocate the treatments among the plots at random.

This randomisation is an essential part of the process. If the treatments exert no effect the observed yields might have occurred in any order, and by making the inference in the proper way we are able to test in the z-distribution without assuming parent normality or the non-existence of fertility differences between plots of the same block. If, of course, the parent is near to normality the test is strengthened. Had we not allocated the treatments at random the use of the z-distribution would not have been valid in the absence of normality (at least approximate) on the part of the parent.

23.42. It is of some importance to make clear the exact hypothesis which is being tested in this approach, since misunderstandings on the point have led to some rather heated controversy. If the treatments are numbered 1 to q, we consider the possible yield on the plot j, k if it received the lth treatment, say $x_{jk\,(l)}$. In actual fact only one of these treatments was carried out; the other values of $x_{jk\,(l)}$ are hypothetical and are based on our conception of what would happen if the treatments were differently distributed. The totality of values $x_{jk\,(l)}$ form our hypothetical population. We are supposing that the observed yields can be expressed as

$$x_{jk\,(l)} = a_j + \xi_{jk\,(l)},$$

where a_j is an effect differing from block to block but constant within blocks, and $\xi_{jk\,(l)}$ is the "individual" plot effect which has a zero mean. The hypothesis we have considered in arriving at the validity of the z-test in conditional inferences is that every treatment affects every plot to the same extent, apart from the block effect a_j. In short, we suppose that $\xi_{jk\,(l)}$ is the same for all l. This is the hypothesis usually tested in data from randomised blocks.

Neyman (1935*a*) proposed an alternative hypothesis, viz. that the mean effects of treatments *over all blocks* were the same, on the ground that we are interested in average treatment effects when testing fertilisers, not the effect on particular plots. The hypothesis here is that $x_{..\,(l)} = x_{..}$, which is not the same as before; and it appears from Neyman's analysis that the z-distribution under randomisation may not hold to such a satisfactory approximation as in the former case. Once again we have to stress the importance of gaining a clear idea of the hypothesis under test.

Example 23.8 (Eden and Yates, 1933; Pitman, 1938)

Eden and Yates considered some data, based on actual experience of heights of wheat shoots, comprising eight classes of four, equivalent to the following measurements :—

Class

1	2	3	4	5	6	7	8
433	455	$487\frac{1}{2}$	$407\frac{1}{2}$	$452\frac{1}{2}$	$257\frac{1}{2}$	$434\frac{1}{2}$	$475\frac{1}{2}$
429	$419\frac{1}{2}$	389	$574\frac{1}{2}$	$436\frac{1}{2}$	$263\frac{1}{2}$	$526\frac{1}{2}$	$473\frac{1}{2}$
383	479	$463\frac{1}{2}$	$477\frac{1}{2}$	415	392	470	$423\frac{1}{2}$
437	$504\frac{1}{2}$	$469\frac{1}{2}$	$452\frac{1}{2}$	418	426	532	$481\frac{1}{2}$

The variances of the eight classes, in units of $\frac{1}{16}$th, are then found to be

7628; 15,702; 22,669; 59,732; 3,666; 90,593; 26,297; 8672.

The quantity K of equation (23.79) is then found to be 0·7577. The quantity $\dfrac{(p-1)(q-1)}{pq-p+2}$ is 0·8077. Thus (23.80) is approximately satisfied and we expect that the z-distribution will be approximately reproduced by the data under random permutations.

This was confirmed by Eden and Yates in a sampling experiment on the data. 1000 sets of permutations were taken and z calculated for each. Agreement with expectation was good.

Example 23.9 (Friedman, 1937)

A good example of data from populations which are probably far from normal is given in Table 23.18, showing the standard deviations of expenditures on various items for six

income-groups. The figures relate to families of wage-earners and lower salaried workers in Minneapolis and St. Paul, U.S.A., in 1935–6.

TABLE 23.18

Standard Deviations of Expenditure on Certain Items of Families in Specified Income Groups.

(Figures in brackets are ranks.)

Category of Expenditure.	Annual Family Income (dollars).						
	750–	1000–	1250–	1500–	1750–	2000–	2250–2500
Housing	100·3 (5)	68·4 (1)	89·5 (3)	77·9 (2)	100·0 (4)	108·2 (6)	184·9 (7)
Household operation	42·2 (1)	44·3 (3)	60·9 (4)	73·9 (6)	43·9 (2)	61·7 (5)	102·3 (7)
Food	71·3 (1)	81·9 (2)	100·7 (7)	86·5 (3)	100·3 (5)	90·7 (4)	100·6 (6)
Clothing	37·6 (1)	60·0 (3)	57·0 (2)	60·8 (4)	71·8 (5)	83·0 (6)	117·1 (7)
Furnishings, etc.	58·3 (2)	52·7 (1)	96·0 (6)	60·4 (3)	104·3 (7)	89·8 (5)	85·8 (4)
Transportation	46·3 (1)	82·2 (2)	129·8 (3)	181·0 (6)	172·3 (5)	164·8 (4)	246·8 (7)
Recreation	19·0 (1)	23·1 (2)	38·7 (3)	45·8 (4)	59·0 (7)	50·7 (5)	55·2 (6)
Personal care	8·3 (1)	8·4 (2)	9·2 (3)	14·3 (6)	10·6 (4)	15·8 (7)	12·5 (5)
Medical care	20·1 (1)	33·5 (2)	60·1 (4)	69·3 (5)	114·3 (7)	45·3 (3)	101·6 (6)
Education	3·2 (1)	4·1 (2)	12·7 (4)	18·9 (5)	8·9 (3)	41·5 (6)	66·3 (7)
Community welfare	4·1 (1)	18·9 (5)	8·5 (2)	12·9 (3)	25·3 (7)	19·9 (6)	16·8 (4)
Vocation	7·7 (1)	11·2 (5)	10·4 (2)	10·9 (4)	10·5 (3)	14·0 (6)	14·4 (7)
Gifts	5·3 (1)	10·9 (2)	11·2 (3)	25·3 (4)	42·3 (5)	48·8 (6)	69·4 (7)
Other	6·0 (5)	5·6 (4)	22·2 (7)	2·5 (2)	6·2 (6)	1·0 (1)	4·0 (3)

In brackets we show the ranks of the figure for different income-groups for each category of expenditure. We wish to know whether the standard deviations for each category differ significantly for the different income levels. On the hypothesis that they do not it is a matter of chance how the ranks fall.

The sums of ranks in each column are :—

23, 36, 53, 57, 70, 70, 83.

The coefficient of concordance (vol. I, p. 411) is then $W = \dfrac{12S}{m^2(n^3 - n)}$, where $m = 14$, $n = 7$ and S is the sum of squares of deviations of sums of ranks from the mean $\dfrac{m(n+1)}{2} = 56$; we find that $S = 2620$ and $W = 0{\cdot}4774$. We may test the significance (vol. I, p. 419) by writing

$$z = \tfrac{1}{2}\log\frac{(m-1)W}{1-W} = 1{\cdot}24$$

$$\nu_1 = (n-1) - \frac{2}{m} = 5\tfrac{6}{7}$$

$$\nu_2 = (m-1)\,\nu_1 = 76\tfrac{1}{7}.$$

The value of z is highly significant, and we conclude that standard deviation is related to size of income—the more money there is to spend, the more variable is the expenditure on particular items.

NOTES AND REFERENCES

The idea of comparing variance between classes with the variance within classes in order to test homogeneity is found as early as Lexis (see footnote on page 119). Modern developments, and particularly the exact test of significance for normal parents, are due mainly to R. A. Fisher. Apart from papers by Irwin (1931 and 1934), connected accounts of the *theory* of variance analysis are hard to find, many points of theoretical interest being scattered among papers which are primarily practical.

For the general theory and applications reference may be made to Fisher's *Statistical Methods* (1925*a*, 1944) and *Design of Experiments* (1935*c*, 1942), to a useful introductory account by Goulden (1939), and to the writings of Yates, particularly his *Design and Analysis of Factorial Experiments* (1937*b*).

On the question of randomisation in preserving the z-distribution see Eden and Yates (1933), Welch (1937, 1938*a*), and Pitman (1938). References to work on ranking are given at the end of Chapter 16.

For work on the distribution of the greatest of a set of variances see Fisher (1929*a*, 1940*a*), Cochran (1941), Stevens (1939*a*), Hartley (1938), and Finney (1941*a*). For further work on the square-root and $\sin^{-1}$ transformations see Cochran (1940*b*), Beall (1942) and Curtiss (1943).

The literature of this subject is now very large. Some further references are given at the end of the next chapter.

EXERCISES

23.1. If x_j $(j = 1 \ldots n)$ are a set of normal independent variates with variances $1/w$, consider the transformation

$$u_k = \sum_{j=1}^{n} l_{kj}\, x_j \sqrt{w_j},$$

where the l's are defined by

$$\begin{array}{ll}
l_{1k} = \sqrt{(w_k/\Sigma w)} & k = 1 \ldots n \\
l_{jk} = \sqrt{\left\{w_j\, w_k \Big/ \left(\sum\limits_{t=1}^{j-1} w_t\right)\left(\sum\limits_{t=1}^{j} w_t\right)\right\}} & \begin{array}{l} j = 2, 3 \ldots n \\ k = 1, 2, \ldots j - 1 \end{array} \\
l_{jk} = -\sum\limits_{t=1}^{j-1} w_t \Big/ \sqrt{\left\{\left(\sum\limits_{t=1}^{j-1} w_t\right)\left(\sum\limits_{t=1}^{j} w_t\right)\right\}} & \begin{array}{l} j = 2, 3, \ldots n \\ k = j \end{array} \\
l_{jk} = 0. & \begin{array}{l} j = 2, 3, \ldots n - 1 \\ k = j + 1, \ldots n \end{array}
\end{array}$$

Show that the l's are orthogonal and hence that

$$\sum_{k=1}^{n} u_k^2 = \sum_{k=1}^{n} w_k\, x_k^2$$

is distributed as χ^2 with n degrees of freedom. Noting that $u_1 = \sum\limits_{k=1}^{n} w_k x_k / \sqrt{\Sigma w}$ is distributed normally with unit variance independently of $u_2 \ldots u_n$, show that

$$\sum_{k=1}^{n} w_k\, (x_k - \bar{x})^2$$

is distributed as χ^2 with $n - 1$ degrees of freedom.

Hence derive the z-test for the analysis of variance with unequal members in a one-way classification.

(Irwin, 1942.)

23.2. Verify the arithmetic in the analysis of variance of Example 23.5.

23.3. Verify the arithmetic in the analysis of variance of Example 23.6.

23.4. In a bivariate table with k rows (different rows corresponding to different values of the x-variate) write

$$h = \frac{1}{\sigma^2} \sum_x n_x (\bar{y}_x - \bar{y})^2$$

$$q = \frac{1}{\sigma^2} \sum_x (n_x s_x^2),$$

where σ^2 is the variance of the y variate, s_x^2 the variance, and n_x the frequency in the row with variate-value x. Thus

$$\frac{\eta_{yx}^2}{1 - \eta_{yx}^2} = \frac{h}{q}$$

and the ratio on the right is the variance-ratio in a one-way classification with unequal numbers.

Show that, for any form of population,

$$E(h) = k - 1 \qquad E(q) = N - k$$

$$\operatorname{var} h = 2(k-1) + (\beta_2 - 3)\left\{\sum_x \frac{1}{n_x} + \frac{1 - 2k}{N}\right\}$$

$$\operatorname{var} q = 2(N-k) + (\beta_2 - 3)\left\{\sum_x \frac{1}{n_x} + N - 2k\right\}$$

$$\operatorname{cov}(h, q) = (\beta_2 - 3)\left\{k - 1 + \frac{k}{N} - \sum_x \frac{1}{n_x}\right\}.$$

Hence, approximately, that

$$E\left(\frac{h}{q}\right) = \frac{E(h)}{E(q)}\left\{1 + \frac{\operatorname{var} q}{E^2(q)} - \frac{\operatorname{cov}(h, q)}{E(h)\,E(q)}\right\}$$

$$E\left(\frac{h}{q}\right)^2 = \frac{E^2(h)}{E^2(q)}\left\{1 + \frac{\operatorname{var} h}{E^2(h)} - \frac{4\operatorname{cov}(h, q)}{E(h)\,E(q)} + \frac{3\operatorname{var} q}{E^2(q)}\right\}.$$

In the case when all rows contain the same frequency

$$\sum\left(\frac{1}{n_x}\right) = \frac{k^2}{N},$$

and then

$$E\left(\frac{h}{q}\right) = \frac{k-1}{N-k}\left\{1 + \frac{2}{N-k}\right\}$$

$$\operatorname{var}\left(\frac{h}{q}\right) = \frac{2(k-1)(N-1)}{(N-k)^3}.$$

Hence show that the mean and variance of the variance-ratio are, to this order, independent of the distribution of y, indicating that the z-test is not very sensitive to deviations from normality.

(E. S. Pearson, 1931*b*. It is rather remarkable that the correlation of h and q, far from disturbing the z-distribution, contributes to its stability.)

CHAPTER 24

THE ANALYSIS OF VARIANCE—(2)

Estimation of Class-differences

24.1. In the previous chapter we considered the analysis of variance mainly as the provider of tests of homogeneity. We have now to examine in more detail the problem of estimating class-effects, assuming that the homogeneity tests have shown them to exist. We discuss in the first instance the case in which there is only one member in each sub-class, and for the sake of simplicity confine ourselves to a two-way classification, though the theory is quite general.

The fundamental hypothesis to be examined is that the data may be expressed in the form

$$x_{jk} = a_j + b_k + \zeta_{jk}, \qquad \ldots \qquad (24.1)$$

where a_j and b_k represent class-effects and ζ is a random normal variate with zero mean. Our analysis of variance will have shown whether this is an acceptable hypothesis, and our present problem is to estimate the unknown values of a's and b's from the observed x's.

24.2. The joint probability of the ζ's is

$$dF \propto \frac{1}{v^{\frac{1}{2}pq}} \exp\left\{-\frac{1}{2v}\Sigma(x_{jk} - a_j - b_k)^2\right\} d\zeta_{11} \ldots d\zeta_{pq}, \qquad \ldots \qquad (24.2)$$

where v is the variance of ζ, and in conformity with the notation used in the previous chapter we have p A-classes and q B-classes. The maximum likelihood estimates of the a's and b's are then those which minimise the sum in curly brackets in (24.2), that is to say, the least-squares solution of the equations (24.1). In the usual way we find

$$\left.\begin{aligned} \sum_{k=1}^{q}(x_{jk} - a_j - b_k) &= 0, \qquad j = 1, \ldots p \\ \sum_{j=1}^{p}(x_{jk} - a_j - b_k) &= 0, \qquad k = 1, \ldots q \end{aligned}\right\} \qquad \ldots \qquad (24.3)$$

which reduce to

$$\left.\begin{aligned} x_{j.} - a_j - b_{.} &= 0 \\ x_{.k} - a_{.} - b_k &= 0 \end{aligned}\right\}. \qquad \ldots \qquad (24.4)$$

Summing the first equation over j, dividing by p, and subtracting from the first, we obtain

$$x_{j.} - x_{..} = a_j - a_{.} \qquad j = 1, \ldots p. \qquad \ldots \qquad (24.5)$$

and similarly

$$x_{.k} - x_{..} = b_k - b_{.} \qquad k = 1, \ldots q. \qquad \ldots \qquad (24.6)$$

In (24.5) there are p equations, but if we sum them all we reach the identity $0 = 0$, so that only $p - 1$ are independent. There is thus an element of indeterminacy which we may remove by supposing that $a_{.} = 0$. Similarly we may take $b_{.} = 0$, and then we have

$$a_j = x_{j.} - x_{..} \qquad j = 1, \ldots p \qquad \ldots \qquad (24.7)$$

$$b_k = x_{.k} - x_{..} \qquad k = 1, \ldots q. \qquad \ldots \qquad (24.8)$$

Our estimate of any class-effect is equal to the deviation of the mean in that class from the total mean.

24.3. Evidently similar equations arise in the general n-way classification. We shall see below that they break down for unequal numbers in subclasses, except in a special case when the numbers are proportionate.

The assumption that a_j and b_k have zero means is not, in effect, a restriction on generality but only a convention. If we prefer it, we may consider the slightly more general hypothesis that ξ has a mean m, in which case we have to minimise

$$\Sigma\,(x_{jk} - a_j - b_k - m)^2. \quad . \quad . \quad . \quad . \quad . \quad (24.9)$$

This will be found to lead back to equations (24.7) and (24.8), with the additional equation for estimating m

$$m = x_{..} \quad . \quad . \quad . \quad . \quad . \quad . \quad (24.10)$$

Or again, if we prefer to absorb m into the a-effects we have

$$\left.\begin{aligned} a_j &= x_{j.} \\ b_k &= x_{.k} - x_{..} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (24.11)$$

the mean of a_j in this case not vanishing. Which form we use is a matter of convenience.

24.4. It is important to notice that the equations of estimation which we have just reached give each a_j and b_k independently of values in other classes. We obtain the same equation for a_j whether we happen to be estimating other a's and b's or not. This property, as we shall see shortly, fails to hold if the numbers in subclasses are disproportionate. The situation is similar to that in which we can determine the constants in a regression line independently of the others if orthogonal polynomials are used, in that each constant is given by a separate equation not containing any of the others. Data of this kind are called *orthogonal*.

The direct comparison of class-means which is possible with orthogonal data can be seen, from general considerations, to be legitimate. In comparing $x_{i.} - x_{..}$ with $x_{j.} - x_{..}$, the estimates of the effects in the ith and jth A-classes, we are in each case averaging over q B-classes with one member in each. The B-classes, therefore, affect each mean to the same extent and do not affect their difference. If there are more members in some subclasses than in others, the means are unequally weighted with different B-effects and the comparison is invalidated.

24.5. Regarding $x_{j.} - x_{..}$ as the estimate of a_j and $x_{.k} - x_{..}$ as the estimate of b_k, we see that the familiar equation

$$\Sigma\,(x_{jk} - x_{..})^2 = \Sigma\,(x_{j.} - x_{..})^2 + \Sigma\,(x_{.k} - x_{..})^2 + \Sigma\,(x_{jk} - x_{j.} - x_{.k} + x_{..})^2 \qquad (24.12)$$

can be regarded as an analysis of the sum of squares on the left, which has $pq - 1$ degrees of freedom, into terms in which there is one degree of freedom for every fitted constant and a residual with $(p - 1)(q - 1)$ degrees of freedom. Every constant fitted reduces the number of degrees of freedom in the residual by unity.

Unequal Numbers in Subclasses

24.6. For a one-way classification we have already considered (**23.7** and **23.8**) the case where the numbers in subclasses are unequal. It was seen that the total sum of squares could be expressed as a sum between classes and a residual which were independently distributed and whose ratio therefore provided a homogeneity test in the usual way.

When we try to extend this result to two-way or generally to n-way classifications, we begin to run into difficulties. We can still find, as shown below, an estimator of v based on $p - 1$ degrees of freedom and differences between A-classes, and one with $q - 1$ d.f. based on differences between B-classes ; but these are no longer independent, and consequently we cannot subtract their sum from the total sum of squares in order to obtain a residual or an interaction term which also provides an unbiassed estimator.

On the other hand, there is now available an independent estimator of v which did not appear in the orthogonal case where only one member was included in each subclass. In fact, since there are several members in any given subclass, we can find an estimator of v based on those members alone ; and we may pool all such to form an estimator with $N - pq$ degrees of freedom, where there are pq subclasses. This estimator will be independent of subclass means and any estimators based on them, and hence provides a "residual" such as we require to carry out homogeneity tests.

24.7. Suppose we have a two-way classification into p A-classes and q B-classes, and let the number of members in the subclass $A_j B_k$ be n_{jk}. Let $\bar{x}_{jk}$ be the mean of these members. We may array the means as

$$\left.\begin{matrix} \bar{x}_{11} & \bar{x}_{12} & \cdot & \cdot & \cdot & \bar{x}_{1q} \\ \bar{x}_{21} & \bar{x}_{22} & \cdot & \cdot & \cdot & \bar{x}_{2q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{x}_{p1} & \bar{x}_{p2} & \cdot & \cdot & \cdot & \bar{x}_{pq} \end{matrix}\right\} \quad . \quad . \quad . \quad . \quad (24.13)$$

Now we may, in the first instance, test for homogeneity by ignoring the differences between A- and B-classification and merely regarding the data as a one-way classification with pq classes. The usual test for homogeneity is then applicable. The sum of squares between means of classes will have $pq - 1$ degrees of freedom, the total $N - 1$ d.f., and the residual $N - 1 - (pq - 1) = N - pq$ d.f. This residual, in fact, is the one mentioned in the previous section, and is based on the pooled sums of squares within the pq classes. The other term based on $pq - 1$ degrees of freedom is the sum

$$\Sigma\, n_{jk} (\bar{x}_{jk} - x_{..})^2$$

and is derivable from the array (24.13).

24.8. To test the effect of A-classification separately we proceed as follows :—

Any $\bar{x}_{jk}$ is the mean of n_{jk} values and, on the usual hypothesis as to normality, will have variance $\frac{v}{n_{jk}}$. If $x_{..}$ is the mean of all N values we have

$$x_{..} = \frac{1}{N} \sum_{j,\,k} n_{jk}\, \bar{x}_{jk} . \quad . \quad . \quad . \quad . \quad . \quad (24.14)$$

Let the marginal *unweighted* means in (24.13) be $\bar{x}_{j.}$, $\bar{x}_{.k}$, so that

$$\left.\begin{aligned} \bar{x}_{j.} &= \frac{1}{q}\sum_k \bar{x}_{jk} \\ \bar{x}_{.k} &= \frac{1}{p}\sum_j \bar{x}_{jk} \end{aligned}\right\} \qquad (24.15)$$

On the hypothesis of homogeneity the variance of $\bar{x}_{j.}$ is given by

$$\frac{v}{q^2}\left(\frac{1}{n_{j1}} + \frac{1}{n_{j2}} + \ldots \frac{1}{n_{jq}}\right) = \frac{v}{N_j}, \text{ say,} \qquad (24.16)$$

where

$$\frac{1}{N_j} = \frac{1}{q^2}\sum_k\left(\frac{1}{n_{jk}}\right). \qquad (24.17)$$

Now let us regard the means $\bar{x}_{j.}$ as the means in p classes whose numbers are N_j, as is legitimate from (24.16). Then writing

$$c = \frac{\Sigma N_j \bar{x}_{j.}}{\Sigma N_j} \qquad (24.18)$$

we have for an unbiassed estimator of v

$$\frac{1}{p-1}\sum_j N_j(\bar{x}_{j.} - c)^2 = \frac{1}{p-1}\left\{\sum_j (N_j \bar{x}^2_{j.}) - c^2 \sum_j N_j\right\}. \qquad (24.19)$$

This estimator has $p - 1$ degrees of freedom and is distributed as χ^2. (This follows from the one-way case except that N_j may not be integral; and its general truth may be established as in Exercise 23.1.) It is independent of the residual with $N - pq$ d.f., and hence the A-effects may be tested separately.

Similarly, if

$$\frac{1}{M_k} = \frac{1}{p^2}\sum_j\left(\frac{1}{n_{jk}}\right), \qquad (24.20)$$

an unbiassed estimator of v is given by

$$\frac{1}{q-1}\left\{\sum_k (M_k \bar{x}^2_{.k}) - d^2 \sum_k M_k\right\}, \qquad (24.21)$$

where

$$d = \frac{\sum_k M_k \bar{x}_{.k}}{\sum_k M_k}, \qquad (24.22)$$

and this also may be compared with the independent estimator based on $N - pq$ d.f.

Example 24.1 (data from Brandt (1933) considered by Yates (1934*a*))

Table 24.1 shows, for a number of breeds of pig, the numbers of each breed, divided into male and female, and the total logarithm of the percentage bacon yielded by the slaughtered carcases. The logarithm has been taken so as to normalise the variate.

TABLE 24.1

Numbers and Logarithm of Percentage Bacon in Breeds of Pigs.

Breed.	Female.		Male.	
	Number.	Log. Percent. Bacon.	Number.	Log. Percent. Bacon.
Hampshire . . .	33	66·55	89	181·04
Duroc Jersey . .	51	98·69	141	281·43
Tamworth . . .	13	25·90	17	34·20
Yorkshire . . .	4	7·62	9	17·58
Berkshire . . .	8	14·64	4	8·20
Poland China . .	15	28·11	32	64·42
Chester White . .	35	66·90	47	90·52
Others	12	23·32	23	46·70
TOTALS . .	171	331·73	362	724·09

The total sum of squares, which is not obtainable from this table as it stands, we quote as 13·0142.

The class-means and reciprocals of class-frequencies are given in Table 24.2.

TABLE 24.2

Class-Means and Reciprocals of Class-Frequencies for the Data of Table 24.1.

Breed.	Female.		Male.		Unweighted Mean of Means.
	Mean.	$1/n_{jk}$	Mean.	$1/n_{jk}$	
Hampshire	2·016,667	0·030,30	2·034,158	0·011,24	2·025,412
Duroc Jersey . . .	1·935,099	0·019,61	1·995,958	0·007,09	1·965,528
Tamworth	1·992,307	0·076,92	2·011,765	0·058,82	2·002,036
Yorkshire	1·905,000	0·250,00	1·953,333	0·111,11	1·929,167
Berkshire	1·830,000	0·125,00	2·050,000	0·250,00	1·940,000
Poland China . . .	1·874,000	0·066,67	2·013,125	0·031,25	1·943,562
Chester White . . .	1·911,429	0·028,57	1·925,958	0·021,28	1·918,694
Others	1·943,333	0·083,33	2·030,434	0·043,48	1·986,884
Unweighted Mean of Means	1·925,979	(Total) 0·680,40	2·001,841	(Total) 0·534,27	1·963,910

Taking first the classification into male and female ($q = 8$), we find, from the relations

$$\frac{1}{N_j} = \frac{1}{q^2} \sum_k \frac{1}{n_{jk}}$$

$$N_1 = \frac{64}{0{\cdot}680{,}40} = 94{\cdot}0623$$

$$N_2 = \frac{64}{0{\cdot}534{,}27} = 119{\cdot}7896.$$

Then, from (24.18)

$$c = \frac{\Sigma N_j \bar{x}_{j.}}{\Sigma N_j} = \frac{(94{\cdot}0623 \times 1{\cdot}925{,}979) + (119{\cdot}7896 \times 2{\cdot}001{,}841)}{94{\cdot}0623 + 119{\cdot}7896}$$

$$= 1{\cdot}968{,}474.$$

Thus our estimate of v, with one degree of freedom

$$= \Sigma (N_j x_{j.}^2) - c^2 (\Sigma N_j)$$
$$= 0{\cdot}3032.$$

Similarly for the eight breed-classes we find an estimate of v with seven degrees of freedom to be $\frac{0{\cdot}6056}{7} = 0{\cdot}0865$.

Considering the 16 subclasses as a one-way classification, we find the following preliminary analysis (the arithmetical details of which we omit) :—

TABLE 24.3

Analysis of Variance of Data in Table 24.1.

Sum of Squares.		d.f.	Quotient.
Between classes	1·2715	15	0·0848
Residual	11·7427	517	0·0227
TOTALS	13·0142	532	

The variance ratio here gives a value of z equal to 0·659, which is significant. Thus the data are not homogeneous.

We now require to decide whether the departure from homogeneity is due to either breed or sex or to a combination of the two. For sex-differences we have found an estimate of v equal to 0·3032 with one d.f. Comparing this with the independent residual from Table 24.3 of 0·0227 with 517 d.f., we find that the effect of sex is significant. Similarly, for breed, the estimate of v is 0·0865 for 7 d.f., which again is significant. We conclude that both breed and sex influence the departure from homogeneity.

It is particularly important to note that since the estimates between breeds and between sex are dependent, we cannot analyse the variance as follows :—

TABLE 24.4

Incorrect Form of Analysis of Variance of Data of Table 24.1.

Sum of Squares.		d.f.	Quotient.
Between sexes	0·3032	1	0·3032
Between breeds	0·6056	7	0·0865
" Interaction "	0·3627	7	0·0518
Residual	11·7427	517	0·0227
TOTALS	13·0142	532	

In fact the term shown as " interaction ", calculated so as to make the sums of squares and degrees of freedom additive in the usual way, is not an unbiassed estimate of v. This is a critical point of difference between the orthogonal and the non-orthogonal case.

24.9. Suppose that the homogeneity test has shown the existence of significant class-effects. As before, we turn to consider the hypothesis that the data can be expressed as the sum of A- and B-effects separately with a random normal residual. Let x_{jkl} be the typical member of the (j, k)th subclass, l varying from 1 to n_{jk}. Our hypothesis is then

$$x_{jkl} = a_j + b_k + \zeta_{jkl}, \qquad (24.23)$$

where ζ is normal with variance v. For convenience we will regard the mean of ζ as absorbed in the coefficients a, so that we may take ζ to have zero mean.

The usual process of estimation of the a's and b's leads to the minimisation of the sum over all N values of

$$\Sigma\,(x_{jkl} - a_j - b_k)^2.$$

Differentiating with respect to a_j and b_k, we find the series of equations

$$\left.\begin{aligned}\sum_k \Sigma'\,(x_{jkl} - a_j - b_k) &= 0, \qquad j = 1 \ldots p\\ \sum_j \Sigma'\,(x_{jkl} - a_j - b_k) &= 0, \qquad k = 1 \ldots q\end{aligned}\right\}, \qquad (24.24)$$

where Σ' denotes summation over the n_{jk} values in a subclass. These equations reduce to

$$\left.\begin{aligned}\sum_k n_{jk}\,a_j + \sum_k n_{jk}\,b_k &= \sum_k n_{jk}\,\bar{x}_{jk}\\ \sum_j n_{jk}\,a_j + \sum_j n_{jk}\,b_k &= \sum_j n_{jk}\,\bar{x}_{jk}\end{aligned}\right\}. \qquad (24.25)$$

Writing $N_{j.}$ for $\sum\limits_k n_{jk}$ and $N_{.k}$ for $\sum\limits_j n_{jk}$, we have

$$N_{j.}\,a_j + \sum_k n_{jk}\,b_k = \sum_k n_{jk}\,\bar{x}_{jk} \qquad j = 1, \ldots p \qquad (24.26)$$

$$\sum_j n_{jk}\,a_j + N_{.k}\,b_k = \sum_j n_{jk}\,\bar{x}_{jk} \qquad k = 1, \ldots q. \qquad (24.27)$$

To which we may add

$$\sum_k b_k = 0. \qquad (24.28)$$

Had we chosen to absorb the mean of ζ into the b's, this last equation would be replaced by $\sum_j a_j = 0$.

When all the n's are equal these equations reduce to the orthogonal case, and each a- or b-coefficient can be independently estimated. In the contrary case the equations have to be solved as they stand.

Example 24.2

Returning to the data of Table 24.1, we find for equations (24.26) and (24.27) the following, the values of the constants required being obtainable from the body or marginal sums of the table itself :—

$$\begin{array}{llllllllllll}
171a_1 & & + 33b_1 & + 51b_2 & + 13b_3 & + 4b_4 & + 8b_5 & + 15b_6 & + 35b_7 & + 12b_8 & = 331\cdot73 \\
& 362a_2 & + 89b_1 & + 141b_2 & + 17b_3 & + 9b_4 & + 4b_5 & + 32b_6 & + 47b_7 & + 23b_8 & = 724\cdot09 \\
33a_1 & + 89a_2 & + 122b_1 & & & & & & & & = 247\cdot59 \\
51a_1 & + 141a_2 & & + 192b_2 & & & & & & & = 380\cdot12 \\
13a_1 & + 17a_2 & & & + 30b_3 & & & & & & = 60\cdot10 \\
4a_1 & + 9a_2 & & & & + 13b_4 & & & & & = 25\cdot20 \\
8a_1 & + 4a_2 & & & & & + 12b_5 & & & & = 22\cdot84 \\
15a_1 & + 32a_2 & & & & & & + 47b_6 & & & = 92\cdot53 \\
35a_1 & + 47a_2 & & & & & & & + 82b_7 & & = 157\cdot42 \\
12a_1 & + 23a_2 & & & & & & & & + 35b_8 & = 70\cdot02
\end{array}$$

To which we may add $a_1 + a_2 = 0$.

The solutions are

$$-a_1 = a_2 = 0\cdot026{,}507 \;;$$

$$b_1 = 2\cdot017{,}259 \;; \quad b_2 = 1\cdot967{,}367 \;; \quad b_3 = 1\cdot999{,}799 \;; \quad b_4 = 1\cdot928{,}267 \;;$$
$$b_5 = 1\cdot912{,}169 \;; \quad b_6 = 1\cdot959{,}136 \;; \quad b_7 = 1\cdot915{,}877 \;; \quad b_8 = 1\cdot992{,}241.$$

These give us the "best" estimates of the mean effects of sex and breed on the hypothesis expressed by (24.23).

The mean of the b's is $1\cdot961{,}514$ which may be taken as an estimate of the mean of ζ, the b-effects then being the differences of the above b-values from this mean.

24.10. Let us now consider the analysis of variance in the non-orthogonal case, when constants have been fitted by least squares in the above-mentioned way.

To make the discussion clearer we will regard the estimation as relating to p constants a_j, related by $\Sigma(a_j) = 0$, q constants b_k, related by $\Sigma(b_k) = 0$, and the mean m. There are thus $p + q - 1$ independent constants which, in effect, provide estimates of the means of subclasses. Whatever these means really are, the residual quotient based on $N - pq$ degrees of freedom gives an unbiassed estimator of v, the common variance. We have now to analyse the remaining sum of squares based on $pq - 1$ d.f.

If the true (population) values of the constants are denoted by α_j, β_k and μ, the sum

$$\Sigma(x_{jkl} - \alpha_j - \beta_k - \mu)^2$$

is distributed as $v\chi^2$ with N degrees of freedom. Developing yet another variation on a familiar theme, we show that the corresponding quantity

$$\Sigma(x_{jkl} - a_j - b_k - m)^2 = \Sigma(x_{jkl} - \alpha_j - \beta_k - \mu)^2 - \Sigma(a_j - \alpha_j)^2 - \Sigma(b_k - \beta_k)^2 - \Sigma(m - \mu)^2 \quad . \quad (24.29)$$

is distributed as $v\chi^2$ with $N - (p + q - 1)$ d.f.

In fact, equations (24.26) and (24.27) show that the estimators a, b (and in our present case m also) are linear in the variables x. We can then find $p + q - 1$ orthogonal normal variables in terms of which they can be expressed. Their sum of squares will be distributed as $v\chi^2$ with $p + q - 1$ degrees of freedom (not some multiple of χ^2 because the mean value must be $p + q - 1$ in virtue of **18.17**). Thus the remaining term $\Sigma (x_{jkl} - a_j - b_k - m)^2$ is distributed as $v\chi^2$ with $N - (p + q - 1)$ degrees of freedom, independently of the portion due to the constants a, b and m.

Furthermore, the actual reduction in sums of squares, equivalent to the sum of the last three terms in (24.29), may be easily determined. Precisely as in the similar problem of evaluating residuals in a regression equation, we have

$$\Sigma (x_{jkl} - a_j - b_k - m)^2 = \Sigma x_{jkl}^2 - \sum_j a_j \sum_{k,l} x_{jkl} - \sum_k b_k \sum_{j,l} x_{jkl} - m \sum_{jkl} x_{jkl} \qquad (24.30)$$

where, of course, summation takes place over all values.

24.11. The total sum of squares is already calculated about the estimated mean m, so that the reduction for the term $\Sigma m^2 = N x_{..}^2$ has already been taken into account. The total sum is then distributed as $v\chi^2$ with $N - 1$ d.f., as we already know. We know further that we can split off the independent residual sum based on $N - pq$ degrees of freedom. This leaves us with a sum based on $pq - 1$ d.f. From the previous section it follows that we can analyse this sum into two parts : (a) the sum of squares due to fitting the constants a_j and b_k, accounting for $p + q - 2$ d.f., and (b) the remainder based on $pq - 1 - (p + q - 2) = (p - 1)(q - 1)$ d.f. This remainder is independent of the sum of squares due to fitting constants and provides an unbiassed estimator of v. If the ratio, as compared with the residual based on $N - pq$ d.f., is significant, the hypothesis of additive effects breaks down. In short, we may regard this quantity as an interaction term.

24.12. One important point to notice in this connection is that the interaction term depends on whether $p + q - 2$ or fewer constants are fitted. In the orthogonal case we can determine an interaction term once and for all, however things stand in regard to the estimation of inter-class effects ; but for non-orthogonal data the number of class-effects estimated affects the interaction term, and if necessary a new significance test has to be applied if further estimates are calculated. The situation is similar to the testing of regression coefficients when orthogonal polynomials are not employed.

Example 24.3

Returning again to the data discussed in Examples 24.1 and 24.2, let us regard the means in all 16 subclasses as simultaneously under estimate. For the reduction in sum of squares due to the constants we find, using the values of a and b found in Example 24.2,—

$$0{\cdot}026{,}507\,(-\,331{\cdot}73 + 724{\cdot}09) + (2{\cdot}017{,}259 \times 247{\cdot}59) + (1{\cdot}967{,}367 \times 380{\cdot}12) \ldots$$

$$- \frac{(1055{\cdot}82)^2}{533} = 1{\cdot}04146.$$

Here, for instance, the sum Σa_1^2 is given by multiplying a_1 by the term $\sum_k x_{1k}$ already found. The last term removes the effect of including the mean among the b's.

The sum of squares between classes was found in Example 24.1 to be 1·2715, based on 15 d.f. We then have

Sum of Squares.		d.f.	Quotient.
Sex and breed (estimation of constants)	1·0415	8	0·1302
Interaction	0·2300	7	0·0329
Between classes	1·2715	15	

Comparing the interaction term 0·0329 (7 d.f.) with the residual 0·0229 (517 d.f.) we see that it is not significant.

If we neglect sex and consider breed alone, we have only to estimate eight constants $b_1 \ldots b_8$ subject to $\Sigma(b) = 0$. The sum of squares for breed alone is given by

$$\frac{1}{122}(247{\cdot}59)^2 + \frac{1}{192}(380{\cdot}12)^2 + \ldots - \frac{1}{533}(1055{\cdot}82)^2 = 0{\cdot}7253.$$

Similarly the sum of squares for sex alone will be found to be 0·4224. We have the following analysis :—

TABLE 24.5

Further Analysis of Variance of Data of Table 24.1.

Sum of Squares.		d.f.	Quotient.
Test for Sex			
Between breed (estimation of constants)	0·7253	7	—
Sex	0·3162	1	0·3162
Sex and breed	1·0415	8	—
Test for Breed			
Between sex (estimation of constants) .	0·4224	1	—
Breed	0·6191	7	0·0884
Sex and breed	1·0415	8	—
Interaction	0·2300	7	0·0329
Between classes	1·2715	15	

Here, for instance, if we test for sex there are seven independent constants for breed and one for sex, the latter being the only one that interests us ; and similarly for breed. On comparison with the residual 0·0227 both sex and breed are found to be significant.

24.13. The reader may perhaps find the various tests of Examples 24.1 and 24.3 confusing, and we accordingly summarise our results for the case of unequal numbers in subclasses.

In every case, except where each subclass contains not more than one member, an estimate of the common variance v may be obtained, with $N - pq$ d.f., by pooling the sums of squares within the pq subclasses. Call this v_1.

Homogeneity may then be tested (a) by considering the pq classes as a single one-way classification and comparing the quotient between means with v_1, or (b) by calculating for either classification separately the estimates based on (24.19) and comparing them with v_1.

If homogeneity is rejected in favour of the additive effect of classes expressed by the usual hypothesis, the sum of squares between all classes based on $pq - 1$ d.f. may be split into independent sums related to the fitting of the constants and to an interaction term. The latter can be compared with v_1 to test for interaction. If this is not significant, alternative tests for effects between A- and between B-classes may be derived by testing the sum of squares attributable to the fitting of the respective constants against v_1. These tests are, in effect, tests of one class neglecting the effect of the other, and may not be accurate if the latter effect is not negligible. It is probably better to fit constants to both classes simultaneously in the first instance.

Proportionate Frequencies

24.14. We have previously spoken of non-orthogonal data as meaning any classification with unequal frequencies in the subclasses, but there is one other case of unequal frequencies for which orthogonality exists, namely the one in which frequencies are proportionate, i.e. there are marginal frequencies l_j, m_k, such that

$$n_{jk} = l_j m_k. \qquad (24.31)$$

Here the means of A-classes are estimates of the individual corresponding a's (though it must not be overlooked that they are based on different numbers of members in margins), and the sum of squares between A-means may be computed in the usual manner appropriate to a one-way classification with unequal numbers. Similarly for B. The interactions may be estimated by subtracting the A- and B-sums from the sum of squares between classes. We leave it to the reader to verify these statements.

Special case of 2×2 . . . Classification

24.15. The foregoing analysis can be extended to the n-way classification, but in the general case the solution of the equations becomes rather complex and the arithmetic a considerable nuisance. Where, however, the classifications are simple dichotomies the problem simplifies to a great extent. For instance, in equations (24.27), if there are only two values of a_j, which we may take to be $+a$ and $-a$, we have

$$N_{.k}\, b_k = \Sigma_j n_{jk}\, \bar{x}_{jk} - n_{1k}\, a + n_{2k}\, a\,.$$

We have selected the a's so that $\Sigma(a) = 0$, which implies that the mean m is amalgamated with the b's. Substituting for the b's in (24.26), we find

$$a\left\{N_{j.} - \Sigma_k n_{1k} \frac{n_{1k} - n_{2k}}{N_{.k}}\right\} = \Sigma_k n_{jk}\, \bar{x}_{jk} - \Sigma_k \frac{n_{1k}}{N_{.k}} \Sigma\, n_j\, \bar{x}_{jk}$$

which reduces to

$$\left(\frac{n_{11}\, n_{12}}{n_{11}+n_{12}} + \frac{n_{21}\, n_{22}}{n_{21}+n_{22}} + \ldots\right) a = \frac{n_{11}\, n_{12}}{n_{11}+n_{12}}(\bar{x}_{11}-\bar{x}_{12}) + \frac{n_{21}\, n_{22}}{n_{21}+n_{22}}(\bar{x}_{21}-\bar{x}_{22}) + \ldots \qquad (24.32)$$

Thus a is the weighted mean of the differences of corresponding B-class means and may be determined direct. So generally for a $2 \times 2 \times 2$ . . . classification. The differences may be tested for homogeneity by the z-test, which in this case reduces to the t-test.

24.16. In view of the relative complexity of the non-orthogonal case, it is natural to wonder whether any serious error would be committed if we regarded the $p \times q$ table of array means as an ordinary two-way table with one member in each class and analysed

the variance accordingly. Evidently such a procedure sacrifices a lot of information about variation in subclasses, but that is not the point. Is the analysis valid ?

The hypothesis on which the analysis is based is equality of variance in subclasses. If the numbers in subclasses are very unequal the means based on them will have very unequal variances, and we expect that the analysis may be misleading. If, however, the numbers are close to equality the analysis will probably be approximately correct.

Example 24.4

Reverting once again to the data considered in earlier examples, we have the following analysis for the variance of the 2×8 table of class-means :—

Sum of Squares.		d.f.	Quotient.
Between sex	0·3032	1	0·3032
Between breed	0·2635	7	0·0376
Residual	0·2387	7	0·0341
TOTALS	0·8054	15	

The sum of squares between sex is the same as before, as it must be for a dichotomy, but the effect of breed is seriously underestimated and would not be judged significant by comparison with the interaction term, which is our residual. The numbers in the breed-classes are, in fact, too different to justify the approximation.

The Missing Plot Technique

24.17. The simplicity of the analysis of variance in the orthogonal case and the economy imported by keeping the number of values as low as possible often leads to the carrying out of experiments with only one member in each subclass. But this has a certain practical danger in that the value in a subclass may be lost through circumstances beyond the experimenter's control. For instance, an animal may die in the course of an experiment, or a crop on a particular plot may be ruined by pest ; or sometimes a record may actually be lost after measurements have been carried out. In such cases we may estimate the missing values and perform a variance-analysis in the following way.

24.18. Consider in the first place a $p \times q$ classification with certain missing values, r in number. We assume as usual that the variate-values are expressible in the form

$$x_{jk} = a_j + b_k + \zeta_{jk} + m, \qquad (24.33)$$

and we know that the "best" estimators of the constants are

$$\left.\begin{aligned} m &= x_{..} \\ a_j &= x_{j.} - x_{..} \\ b_k &= x_{.k} - x_{..} \end{aligned}\right\} \qquad (24.34)$$

The quantities on the right are, however, unknown to us because of the missing values. Suppose that we estimate the constants by minimising

$$\Sigma' (x_{jk} - a_j - b_k - m)^2 \qquad (24.35)$$

where the summation Σ' takes place over *known* values. Our estimators are then determinate and may be written a_j', b_k' and m'.

We will now estimate the missing value on the plot (j, k) by the equation

$$X'_{jk} = a'_j + b'_k + m'. \qquad (24.36)$$

We have

$$\Sigma (x_{jk} - a_j - b_k - m)^2 = \Sigma' (x_{jk} - a_j - b_k - m)^2 + \sum_r (X_{jk} - a_j - b_k - m)^2. \quad (24.37)$$

Let us now consider this as a function to be minimised, involving the unknowns a, b, m and r further unknowns X_{jk}. The equations giving the latter will be obtained by differentiating (24.37) with respect to each X_{jk}, and in fact are typified by

$$X'_{jk} = a'_j + b'_k + m',$$

that is to say, by (24.36). The other constants are given by such equations as

$$\Sigma' (x_{jk} - a'_j - b'_k - m') + \sum_r (X'_{jk} - a'_j - b'_k - m') = 0. \qquad (24.38)$$

The second term vanishes, and hence we obtain the same minimal values for a'_j, b'_k and m' as by minimising (24.35) by itself. Furthermore, the equations of estimation (24.38) may be written

$$\Sigma (x_{jk} - a'_j - b'_k - m') = 0, \qquad (24.39)$$

where the summation takes place over all values, those of the observed x's where known and over the estimated X's where values are missing.

It follows that if we write X_{jk} for the r missing values, ascertain the residual sum of squares, which will be a function of observations and these r unknowns, and minimise it for variation in these unknowns, we shall obtain equations providing estimates of the unknowns equivalent to (24.36). The following example illustrates the method.

Example 24.5 (Yates, 1933*b*)

The following table shows the measurements of intensity of infection of certain potato tubers under eight manurial treatments in ten blocks.

TABLE 24.6

Intensity of Infection of Potato Tubers.

Blocks

Treatments.	1	2	3	4	5	6	7	8	9	10	Totals.
1	3·55	2·29	b	2·00	3·34	3·83	3·86	3·50	2·23	2·91	27·51 $+ b$
2	2·30	4·03	2·54	2·82	3·29	2·93	f	2·55	2·20	2·30	24·96 $+ f$
3	3·96	3·62	3·46	2·50	2·94	3·70	3·82	2·54	3·18	3·69	33·41
4	2·99	3·99	2·90	3·97	4·49	4·70	3·86	h	3·50	3·59	33·99 $+ h$
5	a	3·07	3·49	1·07	3·99	3·48	3·80	3·68	3·24	2·70	28·52 $+ a$
6	2·36	3·47	2·64	3·17	3·26	3·28	g	i	3·07	3·12	24·37 $+ g + i$
7	2·16	2·34	1·96	2·60	3·77	d	3·20	3·47	2·67	3·33	25·50 $+ d$
8	3·16	2·52	2·39	3·68	c	e	3·85	3·36	2·50	4·13	25·59 $+ c + e$
Totals	20·48 $+ a$	25·33	19·38 $+ b$	21·81	25·08 $+ c$	21·92 $+ d + e$	22·39 $+ f + g$	19·10 $+ h + i$	22·59	25·77	223·85 $+ a + b + c + d + e + f + g + h + i$

There are nine missing values in this table, indicated by the letters $a \ldots i$. Omitting purely numerical terms, which are irrelevant for the purposes of minimisation, we have for the total sum of squares,

$$a^2 + b^2 + c^2 + \ldots + i^2 - \tfrac{1}{80}(223{\cdot}85 + a + b + c + \ldots + i)^2;$$

for the sum of squares between blocks,

$$\tfrac{1}{8}\{(20{\cdot}48 + a)^2 + (19{\cdot}38 + b)^2 + \ldots + (19{\cdot}10 + h + i)^2\} - \tfrac{1}{80}(223{\cdot}85 + a + b + c + \ldots + i)^2;$$

and for that between treatments,

$$\tfrac{1}{10}\{(27{\cdot}51 + b)^2 + (24{\cdot}96 + f)^2 + \ldots + (25{\cdot}59 + c + e)^2\} - \tfrac{1}{80}(223{\cdot}85 + a + b + c + \ldots + i)^2.$$

The residual sum of squares is the difference of the first and the sum of the second and third of these expressions. For minimisation we differentiate with respect to $a, b, \ldots i$ in turn. On some arithmetic simplification we find

$$\begin{aligned}
63a + b + c + d + e + f + g + h + i &= 209{\cdot}11\\
a + 63b + c + d + e + f + g + h + i &= 190{\cdot}03\\
a + b + 63c + d - 7e + f + g + h + i &= 231{\cdot}67\\
a + b + c + 63d - 9e + f + g + h + i &= 199{\cdot}35\\
a + b - 7c - 9d + 63e + f + g + h + i &= 200{\cdot}07\\
a + b + c + d + e + 63f - 9g + h + i &= 199{\cdot}73\\
a + b + c + d + e - 9f + 63g + h - 7i &= 195{\cdot}01\\
a + b + c + d + e + f + g + 63h - 9i &= 239{\cdot}07\\
a + b + c + d + e + f - 7g - 9h + 63i &= 162{\cdot}11
\end{aligned}$$

This set of linear equations can, of course, be solved by routine methods, but also by iterative processes as follows :—

The mean of existent values is 3·15. Assume this to be approximately the values of $b, c \ldots i$. Then for a we have, from the first of the above equations—

$$a = \tfrac{1}{63}\{209{\cdot}11 - (8 \times 3{\cdot}15)\} = 2{\cdot}92.$$

Taking this value of a and 3·15 for $c, d \ldots i$, we find for b from the second equation,

$$b = \tfrac{1}{63}\{190{\cdot}03 - (7 \times 3{\cdot}15) - 2{\cdot}92\} = 2{\cdot}62.$$

Similarly, from the third equation,

$$c = \tfrac{1}{63}\{231{\cdot}67 + (2 \times 3{\cdot}15) - 2{\cdot}92 - 2{\cdot}62\} = 3{\cdot}69,$$

and so on. On reaching i we recalculate a from the first equation, using the approximations to the values of the other constants already obtained ; and so on until our values do not alter. In this case only a second approximation is necessary, the values being—

	a	b	c	d	e	f	g	h	i
First Approx. . . .	2·92	2·62	3·69	3·27	3·76	3·26	3·60	3·88	3·22
Second Approx. . .	2·88	2·58	3·73	3·33	3·76	3·32	3·61	3·89	3·22

These are our estimates of missing yields. The treatment means are found to be :—

1	2	3	4	5	6	7	8
3·009	2·828	3·341	3·788	3·140	3·120	2·883	3·308

24.19. The question now arises how we may analyse the variance of data for which missing values have been estimated in this way.

The original data provided a classification with unequal numbers in subclasses and can be analysed by the methods given earlier in the chapter; except that, since no subclass contains more than one member, we cannot find a residual sum of squares within subclasses based on $N - pq$ d.f. ($N - pq$, in fact, is a negative number.) For instance, regarding the data as a one-way classification with $pq - r$ classes, we shall have an analysis of this type :—

Sums of squares	d.f.
Between classes * . .	$p + q - 2$
Residual	$(p-1)(q-1) - r$
Total	$pq - r - 1$

. . . . (24.40)

The effect of the two classifications separately can be dealt with in the manner of Example 24.1.

24.20. Two simplifications are possible. In the first place, since the minimisation of the residual is the same for the original data as for the data completed by estimates of missing values, we can use the latter to compute the residual precisely as for an orthogonal case, which simplifies the arithmetic.

Secondly, it appears that to an adequate approximation we may substitute the estimated values for missing values and analyse the resulting material in the ordinary way as if it were orthogonal. If the proportion of missing values is high this approximation may perhaps break down, and in practice we should probably regard the experiment as ruined. More usually only a few records are missing, and the effect of replacing them by estimates is hardly likely to affect judgments of significance seriously.

Example 24.6

Continuing the analysis of the data of the previous example, we find, for the total sum of squares, 32·1012 with 70 d.f. The analysis of the *completed* data, that is to say the original data plus the estimates of missing values, is as follows :—

Sum of Squares.		d.f.	Quotient.
Between blocks	9·7176	9	1·0797
Between treatments	6·5812	7	0·9402
Residual	17·6902	54	0·3276
TOTALS	33·9890	70	

* It is assumed that no row or column in the two-way classification is entirely empty. If it were, we should have to ignore it and confine attention to the remaining arrays.

Treating the original data as a case of unequal class numbers we find :—

Sum of Squares.		d.f.	Quotient.
Between blocks and treatments	14·4110	16	0·9007
Residual	17·6902	54	0·3276
Totals	32·1012	70	

For blocks only :—

Sum of Squares.		d.f.	Quotient.
Between blocks	8·5690	9	0·9521
Remainder.	5·8420	7	0·8346
Blocks and treatments .	14·4110	16	

For treatments only :—

Sum of Squares.		d.f.	Quotient.
Between treatments	6·2648	7	0·8950
Remainder.	8·1462	9	0·9051
Blocks and treatments .	14·4110	16	

Whether we use the analysis of completed data or the more exact form, we see that differences between blocks and between treatments are significant as judged by the residual variance. The two analyses are, in fact, not very different, and even with as many as nine missing values out of 80 we should not err by substituting estimated values and treating the data as orthogonal.

Relationship with Regression Analysis

24.21. The general n-way classifications to which variance-analysis may be applied are not necessarily determined by a measurable variate. As for contingency tables, rows or columns can be interchanged without affecting the analysis. We can, however, regard a multivariate frequency table as an n-way classification and apply variance-analysis to it ; and just as regression and correlation analysis provide a refinement on contingency analysis because of the arrangement of the classes in order by reference to a variate, so we may to some extent refine the analysis of variance in such a case.

24.22. Consider in the first instance a $p \times q$ table of *frequencies* in the form of a correlation table. We will suppose the A-classification to be according to the variate x

and the B-classification according to y. Let us now consider the hypothesis that the data emanate from a normal bivariate population with zero correlation (or, somewhat more generally, that for any given y the x's are distributed normally with the same mean and variance). We can then regard the data as a *one-way* classification according to y with unequal frequencies and analyse the variance in the usual form :—

Sum of Squares.		d.f.	Quotient.
Between classes . . .	$\sum_{j=1}^{q} n_j (\bar{x}_j - \bar{x})^2$	$q - 1$	$\frac{N \eta^2 \operatorname{var} x}{q - 1}$
Residual	$\Sigma (x_{ij} - \bar{x}_j)^2$	$N - q$	$\frac{N (1 - \eta^2) \operatorname{var} x}{N - q}$
TOTALS . . .	$N \operatorname{var} x$	$N - 1$	

Here $\bar{x}_j$ is the mean of n_j x-values in the jth y-class, $\bar{x}$ is the mean of all N values, x_{ij} is the variate-value in the ith x-class and jth y-class, and there are q y-classes. The quotients are expressible in terms of the correlation ratio of x on y, viz. η_{xy} (cf. **14.23.** vol. I, p. 351).

Now, on our hypothesis, the sums of squares between classes and the residual are independently distributed in the Type III form, and hence the variance ratio

$$\frac{\eta^2}{q - 1} \frac{N - q}{1 - \eta^2} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (24.41)$$

can be tested in Fisher's distribution with $\nu_1 = q - 1$, $\nu_2 = N - q$. This is the test we gave in **14.25** (vol. I, p. 353) and it is reached by an argument of essentially the same kind.

24.23. Now suppose that our $p \times q$ table is normal but correlated ; or, somewhat more generally, that the values in arrays of constant y are normally distributed with the same variance but with means which vary linearly with y, say

$$m_j = m + by_j. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (24.42)$$

Then our data can be represented by the form

$$x_{ij} = m + by_j + \zeta_{ij}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (24.43)$$

where the ζ's are distributed normally with zero mean and the same variance v. Apart from the constant m, the only unknown here is the constant b. Our least-squares estimates (measuring from the means of x and y) now lead to the familiar form for the regression coefficient

$$b = \frac{\Sigma\, x_{ij}\, y_j}{\Sigma\, y_j^2}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (24.44)$$

where summation takes place over all values observed. This is, of course, equivalent to

$$b = \frac{\operatorname{cov}(x, y)}{\operatorname{var} y}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (24.45)$$

Further, the reduction in sum of squares attributable to fitting the constant b is

$$Nb \operatorname{cov}(x, y) = \frac{N \operatorname{cov}^2(x, y)}{\operatorname{var} y} = N r^2 \operatorname{var} x, \qquad (24.46)$$

where r is the correlation coefficient of the sample.

Our analysis of variance may then be written—

TABLE 24.7

Analysis of Variance of a Correlation Table

Sum of Squares.		d.f.	Quotient.
Regression constant b	$Nr^2 \operatorname{var} x$	1	$Nr^2 \operatorname{var} x$
Between classes (after regression is eliminated)	$N(\eta^2 - r^2) \operatorname{var} x$	$q - 2$	$N \dfrac{\eta^2 - r^2}{q - 2} \operatorname{var} x$
Residual	$N(1 - \eta^2) \operatorname{var} x$	$N - q$	$N \dfrac{1 - \eta^2}{N - q} \operatorname{var} x$
TOTALS	$N \operatorname{var} x$	$N - 1$	

This analysis gives us a test of the significance of the correlation coefficient in samples from an uncorrelated population and also of linearity of regression.

In fact, if the parent correlation is zero, the parent value of b is zero and the quotient due to b is independent of the sum of the other items in the analysis. Thus the ratio

$$\frac{Nr^2 \operatorname{var} x}{N(1 - r^2) \operatorname{var} x} = \frac{r^2}{1 - r^2} \qquad (24.47)$$

is distributed in Fisher's form with $\nu_1 = 1$, $\nu_2 = N - 2$. This is equivalent to saying that

$$\sqrt{\frac{r^2(N - 2)}{1 - r^2}} \qquad (24.48)$$

is distributed in "Student's" form with $N - 2$ d.f., which brings us back by a different route to the test given in **14.15** (vol. I, p. 342).

24.24. Secondly, if we assume that the parent correlation is not zero but the regression is linear, the sum of squares between classes after regression is eliminated is independent of the residual in Table 24.7, and hence the ratio

$$\frac{N \operatorname{var} x \dfrac{\eta^2 - r^2}{q - 2}}{N \operatorname{var} x \dfrac{1 - \eta^2}{N - q}} = \frac{\eta^2 - r^2}{q - 2} \, \frac{N - q}{1 - \eta^2} \qquad (24.49)$$

is distributed in Fisher's form with $\nu_1 = q - 2$, $\nu_2 = N - q$. This test (due to Fisher himself) gives a test of linearity of regression in the normal case.

It should be noticed that this test is only approximate if the classification is one of a normal population with broad groupings. If correlation exists, the distribution of a bivariate normal sample in an array of finite width is not exactly normal, being the sum

of a number of normal distributions with slightly different means. Unless the grouping is very coarse, this is not likely to invalidate tests of significance in practice.

24.25. Consider now the general regression formula for p variates,—

$$x_1 = b_2 x_2 + b_3 x_3 + \ldots + b_p x_p. \qquad (24.50)$$

If we assume that the residuals $x_1 - \sum_{j=2}^{p} b_j x_j$ (say x) are distributed normally with constant variance, our least-squares estimates of the regression coefficients are those given by the usual theory, and the fitting of $(p - 1)$ constants reduces the sum of squares by $N \operatorname{var} x\, R^2$, where R is the multiple correlation coefficient (cf. **15.16,** vol. I, p. 380). We then have the analysis—

Sum of Squares.		d.f.	Quotient.
Between classes (regression constants)	$N \operatorname{var} x\, R^2$	$p - 1$	$\frac{R^2}{p-1} N \operatorname{var} x$
Residual	$N \operatorname{var} x\,(1 - R^2)$	$N - p$	$\frac{1 - R^2}{N - p} N \operatorname{var} x$
TOTALS	$N \operatorname{var} x$	$N - 1$	

If the regression is in fact linear of type (24.50), the residual quotient is independent of that due to fitting regression constants, and the hypothesis may be tested by means of the ratio

$$\frac{R^2}{p-1}\,\frac{N-p}{1-R^2}\,. \qquad (24.51)$$

which is distributed in Fisher's form with $\nu_1 = p - 1$, $\nu_2 = N - p$. This brings us to the distribution of R^2 given in **15.20.**

24.26. It is to be observed that in (24.50) we may choose the variates $x_2 \ldots x_p$ as we please. In particular, we can take them to be polynomials of a single variate. From this point of view the analysis of variance links up with the theory of regression analysis, given in Chapter 22. If the polynomials are orthogonal we can fit the constants b one at a time, the fitting of any constant leaving unchanged the previous determination of those of lower orders. The reduction in sum of squares for each constant can be separately ascertained and corresponds to the loss of a further degree of freedom; and at any stage we may test the residual variance to see whether any particular term is worth while in the sense that it makes a significant contribution to the total variance. The exact test, of course, depends on the usual assumptions of normality.

24.27. The reader is now in a position to see a number of statistical topics which on the surface appear to be distinct as parts of a single theory. Regression analysis, with its subsidiary of correlation analysis, proceeds by the successive fitting of constants by least-squares. For the normal case this is equivalent to estimation by maximum likelihood. Partial and multiple regression, together with curvilinear regression, can all be subsumed

under this central idea. The fitting of each constant splits off a separate contribution to the total variance which, under certain hypotheses, is independent of the others. Variance-analysis proceeds in much the same way, but is more general in the sense that it can deal with the classification of values, however determined. Our various exact tests of significance of homogeneity in variance, of linearity of regression, of significance of correlations in uncorrelated material, of the difference of two means where variances are equal, of the correlation ratios, of the multiple correlation coefficient—all derive ultimately from Fisher's distribution of the variance-ratio in the normal case.

The Analysis of Covariance

24.28. Suppose that we have a one-way classification, possibly with unequal numbers, and that in each class the members present values not of a single variate, such as we have considered up to now, but pairs of variate-values typified by x_{ij}, y_{ij}, j referring as usual to class and i to the number within the class. By the ordinary methods of variance-analysis we can discuss the effect of classification either on the x-variate or on the y-variate; but there also arises for consideration the effect of class-membership on the covariation of x and y. This leads us to an extension of the analysis of variance to that of covariance.

24.29. By an easy extension of the results for a single variate we have, analogously to

$$\sum_{i,j}(x_{ij}-x_{..})^2=\sum_{i,j}(x_{ij}-x_{.j})^2+\sum_{j}n_j(x_{.j}-x_{..})^2$$

the equation in product terms

$$\sum_{i,j}(x_{ij}-x_{..})(y_{ij}-y_{..})=\sum_{i,j}(x_{ij}-x_{.j})(y_{ij}-y_{.j})+\sum_{j}n_j(x_{.j}-x_{..})(y_{.j}-y_{..}) \quad (24.52)$$

If we consider the whole sample as homogeneous the correlation between x and y is given by

$$r=\frac{\Sigma(x_{ij}-x_{..})(y_{ij}-y_{..})}{\sqrt{\{\Sigma(x_{ij}-x_{..})^2\,\Sigma(y_{ij}-y_{..})^2\}}}. \quad . \quad . \quad . \quad (24.53)$$

We have also the correlation between means of classes

$$r=\frac{\Sigma(x_{.j}-x_{..})(y_{.j}-y_{..})}{\sqrt{\{\Sigma(x_{.j}-x_{..})^2\,\Sigma(y_{.j}-y_{..})^2\}}} \quad . \quad . \quad . \quad (24.54)$$

and may calculate a correlation of residuals within classes

$$r=\frac{\Sigma(x_{ij}-x_{.j})(y_{ij}-y_{.j})}{\sqrt{\{\Sigma(x_{ij}-x_{.j})^2\,\Sigma(y_{ij}-y_{.j})^2\}}}. \quad . \quad . \quad . \quad (24.55)$$

24.30. If there is heterogeneity present we should expect these correlations to differ; and similarly for the three kinds of regression of y on x, such as

$$b=\frac{\Sigma(x_{ij}-x_{..})(y_{ij}-y_{..})}{\Sigma(x_{ij}-x_{..})^2}. \quad . \quad . \quad . \quad . \quad (24.56)$$

The three correlations of (24.53)–(24.55) are, however, not additive, like sums of squares; nor are the regressions corresponding. The covariances expressed by (24.52) are additive, but there is no simple test, such as exists for variance-ratios, to determine the significance of differences or ratios of covariances. Covariance analysis, however, is not primarily designed to test independence, but to examine whether there is any variation according

to class between the regressions of y on x within and between classes. Let us suppose that there is some linear relation of the form

$$Y - \mu_y = \beta(X - \mu_x). \qquad (24.57)$$

Following the notation of E. S. Pearson, we write

$$\left.\begin{aligned} C_{11j} &= \sum_i (x_{ij} - x_{.j})^2 \\ C_{22j} &= \sum_i (y_{ij} - y_{.j})^2 \\ C_{12j} &= \sum_i (x_{ij} - x_{.j})(y_{ij} - y_{.j}) \end{aligned}\right\} \qquad (24.58)$$

$$\left.\begin{aligned} C_{11a} &= \sum_j C_{11j} \\ C_{22a} &= \sum_j C_{22j} \\ C_{12a} &= \sum_j C_{12j} \end{aligned}\right\} \qquad (24.59)$$

$$\left.\begin{aligned} C_{11m} &= \sum_j n_j (x_{.j} - x_{..})^2 \\ C_{22m} &= \sum_j n_j (y_{.j} - y_{..})^2 \\ C_{12m} &= \sum_j n_j (x_{.j} - x_{..})(y_{.j} - y_{..}) \end{aligned}\right\} \qquad (24.60)$$

and C_{110}, C_{220}, C_{120} for the corresponding total sums of squares and products. We may then exhibit the composition of the total sums of squares and products in the form of Table 24.8. The arithmetic of the analysis follows that of ordinary variance-analysis. We shall give an example presently.

TABLE 24.8

Analysis of Variance and Covariance for One-Way Classification—Sums of Squares and Products and Regression Coefficients.

Variation.	d.f.	Sum of Squares. x-variate.	Sum of Squares. y-variate.	Sum of Products. xy.	Regression Coefficients.
Within jth group	$n_j - 1$	C_{11j}	C_{22j}	C_{12j}	$b_j = \dfrac{C_{12j}}{C_{11j}}$
Within groups .	$N - p$	C_{11a}	C_{22a}	C_{12a}	$b_a = \dfrac{C_{12a}}{C_{11a}}$
Between groups	$p - 1$	C_{11m}	C_{22m}	C_{12m}	$b_m = \dfrac{C_{12m}}{C_{11m}}$
TOTALS .	$N - 1$	C_{110}	C_{220}	C_{120}	$b_0 = \dfrac{C_{120}}{C_{110}}$

We now suppose that, apart from the regression effects represented by (24.57), the variation of x is normal with constant variance v. We can then compile various estimates of v from the residual variation after the effect of fitting regression constants has been

removed. For instance, within classes we have for the estimator of v, with $N - 2p$ degrees of freedom,

$$\frac{1}{N-2p}\left[\sum_{i,j} \{y_{ij} - y_{.j} - b_j(x_{ij} - x_{.j})\}^2\right]$$
$$= \frac{1}{N-2p} \sum_j (C_{22j} - b_j C_{12j})$$
$$= \frac{1}{N-2p} S_1, \text{ say.}$$

The number of degrees of freedom follows from the fact that we have fitted a mean and a regression coefficient to each of p classes, making a reduction of $2p$ in all. We then obtain Table 24.9 :—

TABLE 24.9

Analysis of Covariance for One-Way Classification with Linear Regressions.

Variation due to	d.f.	Sum of Squares.
Deviations from linear regressions within classes	$N - 2p$	$\sum_{i,j} \{y_{ij} - y_{.j} - b_j(x_{ij} - x_{.j})\}^2 = \sum_j (C_{22j} - b_j C_{12j}) \quad = S_1$
Differences among regressions . .	$p - 1$	$\sum_{i,j} (b_j - b_a)^2 (x_{ij} - x_{.j})^2 = \sum_j (b_j C_{12j}) - b_a C_{12a} \quad = S_2$
Deviations within classes from linear regression b_a	$N - p - 1$	$\sum_{i,j} \{y_{ij} - y_{.j} - b_a(x_{ij} - x_{.j})\}^2 = C_{22a} - b_a C_{12a} \quad = S_1 + S_2$
Deviations between classes from linear regression b_m	$p - 2$	$\sum_j n_j \{y_{.j} - y_{..} - b_m(x_{.j} - x_{..})\}^2 = C_{22m} - b_m C_{12m} \quad = S_3$
Differences between b_a and b_m .	1	$\sum_{i,j} \{(b_a - b_m)(x_{ij} - x_{.j}) + (b_m - b_0)(x_{ij} - x_{..})\}^2 = (b_a - b_m)^2 \frac{C_{11a} C_{11m}}{C_{110}} \quad = S_4$
Total deviation from linear regression b_0	$N - 2$	$\sum_{i,j} \{y_{ij} - y_{..} - b_0(x_{ij} - x_{..})\}^2 = C_{220} - b_0 C_{120} \quad = S_1 + S_2 + S_3 + S_4$

The reader will probably find it useful to check the expressions in the third column of Table 24.9 and to examine how the sum of squares of deviations from the regression line of the whole is analysed into the constituent items.

24.31. Suppose now that we wish to test whether the relationship between x and y can be represented by the formula (24.57), and that there is no material class-effect present. Then S_1 of Table 24.9 should be an unbiassed estimator of $(N - 2p)\,v$ and should be independent of the residual estimator $S_2 + S_3 + S_4$, which has $2p - 2$ d.f. We may therefore test the hypothesis by the ratio

$$\frac{S_1}{N-2p}\cdot\frac{2p-2}{S_2+S_3+S_4}, \qquad \nu_1 = N - 2p, \qquad \nu_2 = 2p - 2. \qquad (24.61)$$

If this variance ratio is insignificant we consider next whether the regressions differ in the p classes. For this purpose we compare the estimator derived from S_2 with that based on S_1; i.e. the ratio

$$\frac{S_2}{p-1}\cdot\frac{N-2p}{S_1}, \qquad \nu_1 = p - 1, \qquad \nu_2 = N - 2p \qquad (24.62)$$

will be significant if differences are to be regarded as real.

If this ratio is not significant, S_1 and S_2 may be pooled. Comparison of their sum with S_3 will afford a test whether the relation between group means is linear. The ratio for this purpose is

$$\frac{S_1+S_2}{N-p-1}\cdot\frac{p-2}{S_3}, \qquad \nu_1 = N - p - 1, \qquad \nu_2 = p - 2. \qquad (24.63)$$

Finally, even if this ratio is not significant, it does not follow that the common regression within groups is the same as the regression of the means of groups. To test this point we consider the ratio

$$\frac{S_1+S_2}{N-p-1}\cdot\frac{1}{S_4}, \qquad \nu_1 = N - p - 1, \qquad \nu_2 = 1. \qquad (24.64)$$

Example 24.7

A number of recruits are given a preliminary test to ascertain their suitability for a certain course of training. At the end of the training course they undergo a proficiency test. The marks for three groups of recruits from three different towns are—

Group 1	Preliminary:	45, 50, 56, 58, 59, 60, 62, 64, 65, 75
	Proficiency:	46, 60, 52, 46, 48, 50, 55, 63, 58, 64
Group 2	Preliminary:	44, 49, 52, 52, 58, 59, 60, 62, 63, 63, 66, 69, 70, 72, 73
	Proficiency:	48, 55, 45, 60, 65, 64, 69, 71, 77, 70, 75, 80, 72, 75, 81
Group 3	Preliminary:	47, 52, 59, 60, 63, 66, 68, 69, 74, 76
	Proficiency:	43, 56, 51, 72, 60, 61, 55, 74, 72, 80.

We are interested here in the efficiency of the preliminary test as a predictor of the proficiency test. We therefore consider the regression of the marks obtained in the latter (y) on those obtained in the former (x). We are, however, also very much interested in the question whether the regressions are the same, apart from purely sampling effects, in the three groups. Such a matter would naturally arise, for instance, if we were thinking

of applying the same rejection standards in preliminary tests to all recruits, irrespective of their town of origin.

Our scores are given to the nearest unit, and hence the variates are discontinuous. We will neglect this effect and assume that the scores are distributed approximately normally.

About origin $x = y = 50$ the sums of squares and cross-products are :—

	n.	$\Sigma(x)$.	$\Sigma(y)$.	$\Sigma(x^2)$.	$\Sigma(y^2)$.	$\Sigma(xy)$.
Group 1	10	94	42	1496	594	694
Group 2	15	162	257	2802	6101	3989
Group 3	10	134	124	2556	2776	2422

We can then calculate the quantities C. For instance,

$$C_{111} = 1496 - 94\frac{94}{10} = 612 \cdot 4$$

$$C_{121} = 694 - 42\frac{94}{10} = 299 \cdot 2$$

$$C_{11a} = C_{111} + C_{112} + C_{113}, \text{ etc.}$$

We find the following table in the form of Table 24.8 :—

TABLE 24.10

Analysis of Variance and Covariance for Data of Example 24.7—Sums of Squares and Products and Regressions

Variation.	d.f.	Sum of Squares. x^2.	Sum of Squares. y^2.	Sum of Products. xy.	Regressions.
Within first group	9	$C_{111} = 612 \cdot 4$	$C_{221} = 417 \cdot 6$	$C_{121} = 299 \cdot 2$	$b_1 = 0 \cdot 4886$
,, second group	14	$C_{112} = 1052 \cdot 4$	$C_{222} = 1697 \cdot 73$	$C_{122} = 1213 \cdot 4$	$b_2 = 1 \cdot 1530$
,, third group	9	$C_{113} = 760 \cdot 4$	$C_{223} = 1238 \cdot 4$	$C_{123} = 760 \cdot 4$	$b_3 = 1 \cdot 0000$
Within groups . .	32	$C_{11a} = 2425 \cdot 2$	$C_{22a} = 3353 \cdot 73$	$C_{12a} = 2273 \cdot 0$	$b_a = 0 \cdot 9372$
Between groups .	2	$C_{11m} = 83 \cdot 09$	$C_{22m} = 1005 \cdot 01$	$C_{12m} = 118 \cdot 57$	$b_m = 1 \cdot 4270$
TOTALS .	34	$C_{110} = 2508 \cdot 29$	$C_{220} = 4358 \cdot 74$	$C_{120} = 2391 \cdot 57$	$b_0 = 0 \cdot 9535$

A comparison of the three regressions within groups indicates some heterogeneity. It looks as if the preliminary test is not such a good predictor for the first group as for the others. We may proceed to test the reality of this effect by constructing Table 24.11 on the lines of Table 24.9. For instance,

$$S_1 = \sum_j (C_{22j} - C_{12j}\, b_j) = (417 \cdot 6 - 299 \cdot 2 \times 0 \cdot 4886) + (\text{two similar terms})$$

$$= 1048 \cdot 1.$$

We find—

TABLE 24.11

Analysis of Covariance of Data of Example 24.7—Linear Regressions.

Variation.	d.f.	Sums S.	Quotient.
Deviations from regressions b_j	29	$S_1 = 1048{\cdot}1$	36·1
Differences b_j	2	$S_2 = 175{\cdot}4$	87·7
Deviations from b_a	31	$S_1 + S_2 = 1223{\cdot}5$	39·5
Deviations of groups from b_m	1	$S_3 = 835{\cdot}6$	835·6
Difference between b_a and b_m	1	$S_4 = 19{\cdot}3$	19·3
TOTALS	33	$S_1 + S_2 + S_3 + S_4 = 2078{\cdot}4$	

A comparison of the quotient 36·1 (29 d.f.) with the quotient of the remaining items, 257·6 (4 d.f.) indicates that there are real differences between classes. A single regression equation will not represent all three class-relations. A comparison of the deviations from regressions, 36·1 (29 d.f.), with the differences of regressions among themselves, 87·7 (2 d.f.), does not reject the hypothesis of equality of regressions within groups. We therefore compare the deviations from b_a, 39·5 (31 d.f.), with the deviations of groups from b_m, 835·6 (1 d.f.). This is significant, suggesting that the hypothesis of linearity of regression of group-means should be rejected.

The general result is to confirm our suspicion of heterogeneity. The correlation coefficients between x and y are—

Within first group	0·592
,, second group	0·908
,, third group	0·784
Within groups	0·797
Between groups	0·410
Total	0·722

Again the deviations between groups stand out as indicating heterogeneity.

24.32. The analysis of covariance may be extended to the case where there is more than one independent variate. The regression coefficients are found in the usual way, and the sums of squares after regressions have been removed can be found and compared on the usual hypotheses. Suppose, for instance, there are two independent variates and a classification giving an analysis between classes and residual. We may represent the analysis thus :—

	d.f.	Sum of Squares.			Sum of Products.		
		x_1^2	x_2^2	y^2	$x_1 x_2$	yx_1	yx_2
Between classes	n	A	B	C	P	Q	R
Residual	n'	A'	B'	C'	P'	Q'	R'
TOTALS	n''	A''	B''	C''	P''	Q''	R''

Our regressions are then—

	b_1	b_2
Between classes	$\dfrac{BQ - PR}{AB - P^2}$	$\dfrac{AR - PQ}{AB - P^2}$
Residual	$\dfrac{B'Q' - P'R'}{A'B' - P'^2}$	$\dfrac{A'R' - P'Q'}{A'B' - P'^2}$
TOTALS . . .	$\dfrac{B''Q'' - P''R''}{A''B'' - P''^2}$	$\dfrac{A''R'' - P''Q''}{A''B'' - P''^2}$

The sums of squares C can then be reduced by eliminating regressions, i.e. by subtracting $Qb_1 + Rb_2$, giving

$$C - \frac{BQ^2 - PQR}{AB - P^2} - \frac{AR^2 - PQR}{AB - P^2}$$

$$= \frac{ABC - AR^2 - BQ^2 - CP^2 + 2PQR}{AB - P^2} \quad . \quad . \quad . \quad (24.65)$$

This and the analogous quantities with primes give independent estimators of the variance of the residual element, and a comparison to test homogeneity may be made in the usual way.

24.33. In a case such as that of Example 24.7 it is evident that a comparison of y-means between groups is affected by what we know about the x-values. If we know nothing about the latter, comparison of the y's is a univariate problem and can be treated by the methods already discussed, the difference of means, for example, being tested by the use of standard errors or the t-test. But suppose that our x's themselves are found to be different between groups and that there is significant correlation between x and y. Then it is possible that the relation, if any, between y's in different groups is not, so to speak, an inherent quality of the variation of y, but is merely a reflection of their dependence on the x's, which happen to exhibit significant differences. In Example 24.7, differences in proficiency between groups may be due simply to differences of ability which were present before the training began and, if so, should be shown by differences between groups in the preliminary scores. We should not then be able to conclude from proficiency scores alone that training in one group had a more marked effect than in another. The differences were there before the training was applied.

24.34. If, then, we require to consider the effects of training alone on the groups, we may "correct" the y-values by deducting the estimates

$$Y_{ij} = y_{..} + b_0 (x_{ij} - x_{..}) \quad . \quad . \quad . \quad . \quad . \quad (24.66)$$

or other more general regression equations. This, so to speak, allows for differences due to variations of the x-variate.

Assuming that one linear regression equation adequately describes the relationship between y and x, so that the corrected values are

$$y_{ij} - Y_{ij} = y_{ij} - y_{..} - b_0 (x_{ij} - x_{..}), \qquad (24.67)$$

we see that the difference of the corrected means of two classes $y_{.j}$ and $y_{.k}$ is

$$y_{.j} - y_{.k} - b_0 (x_{.j} - x_{.k}). \qquad (24.68)$$

This may be regarded as the sum of two parts which are independent. The estimated variance of the first part, $y_{.j} - y_{.k}$, is $\frac{2s^2}{q}$, where s^2 is the mean-square of the residual after correcting for regression and the means of $y_{.j}$ and $y_{.k}$ are both based on q members. Similarly the variance of b is $\frac{s^2}{A}$, where A is the sum of squares of the x-variate entering into the residual row of the analysis. Regarding the x's as fixed from sample to sample, so that our inference is conditional, we see that the variance of the difference (24.68) is given by

$$s^2 \left\{ \frac{2}{q} + \frac{(x_{.j} - x_{.k})^2}{A} \right\}. \qquad (24.69)$$

The ratio of the difference to the square root of this expression is distributed as "Student's" t, with degrees of freedom one fewer in number than those of the original residual.

24.35. Similarly, if we have two independent variables x_1 and x_2, the corrected difference of y-means is

$$y_{.j} - y_{.k} - \{b_1 (x_{1j} - x_{1k}) + b_2 (x_{2j} - x_{2k})\} \qquad (24.70)$$

where temporarily we write x_{1j} for the mean of the variate x_1 in the jth class, and so on. The variance of the part in curly brackets may be derived by considering the variance of the general expression $\lambda b_1 + \mu b_2$. From the equations for b_1 and b_2 we have

$$\left. \begin{aligned} b_1 &= \frac{B\,\Sigma(yx_1) - P\,\Sigma(yx_2)}{AB - P^2} \\ b_2 &= \frac{-P\,\Sigma(yx_1) + A\,\Sigma(yx_2)}{AB - P^2} \end{aligned} \right\} \qquad (24.71)$$

where, as in **24.32**, A and B are the sums of squares for x_1, x_2, and P is the cross-product. Thus the coefficient of any y in $\lambda b_1 + \mu b_2$ is

$$\frac{(\lambda B - \mu P)\, x_1 + (\mu A - \lambda P)\, x_2}{AB - P^2}.$$

Since the y's are independent the estimated variance of $\lambda b_1 + \mu b_2$ is

$$\frac{s^2}{(AB - P^2)^2} \{ A(\lambda B - \mu P)^2 + 2P(\lambda B - \mu P)(\mu A - \lambda P) + B(\mu A - \lambda P)^2 \}$$

$$= \frac{\lambda^2 B - 2\lambda\mu P + \mu^2 A}{AB - P^2} s^2. \qquad (24.72)$$

Thus for the estimated variance of the corrected difference (24.70) we have

$$s^2 \left\{ \frac{2}{q} + \frac{\lambda^2 B - 2\lambda\mu P + \mu^2 A}{AB - P^2} \right\} \qquad (24.73)$$

where $\lambda = x_{1j} - x_{1k}$ and $\mu = x_{2j} - x_{2k}$. As usual, the difference divided by the square root of this quantity may be tested in the t-distribution.

24.36. Our account of the analysis of variance and covariance has not attempted to cover all the applications of the method in particular directions. We have concentrated so far as possible on the fundamental ideas and the broad lines of analysis to which they lead. Some further developments will be given in later chapters, but we must refer the reader who requires a complete acquaintance with the subject to the references given at the end of this chapter and the preceding. We will conclude our exposition with three final comments.

(*a*) Part of our hypothesis throughout has been that the residual element ζ has constant variance from one subclass to another. In Chapter 26 we shall discuss methods of testing homogeneity in residual variance. For completeness we might perhaps have anticipated some of these tests in the present chapter, at least to the extent of exemplifying their use. We have not done so mainly for reasons of economy in space ; but the omission of mention of the point in foregoing examples should not lead the reader to overlook (as many writers do overlook) the necessity for testing variance-homogeneity where possible, if it is required as part of the hypothesis.

(*b*) In the majority of our examples we have proceeded at once to analyses of variance or covariance without dwelling on points which would require attention in any practical inquiry. For instance, since the primary function of many variance-analyses is to test the homogeneity of a set of class-means, the first stage would be to compute those means and examine whether they suggest any lack of homogeneity on intuitive grounds. Again, if heterogeneity is established, consideration of the means themselves, or of the primary data, will sometimes show how it arises. The student must never lose sight of his primary material.

(*c*) Elaborating this point to some extent, we would emphasise that the analysis of variance, like other statistical techniques, is not a mill which will grind out results automatically without care or forethought on the part of the operator. It is a rather delicate instrument which can be called into play when precision is needed, but requires skill as well as enthusiasm to apply to the best advantage. The reader who roves among the literature of the subject will sometimes find elaborate analyses applied to data in order to prove something which was almost obvious from careful inspection right from the start ; or he will find results stated without qualification as " significant " without any attempt at critical appreciation. This is not the occasion to deliver a homily on the necessity for self-discipline in the use of advanced theoretical techniques, but the analysis of variance would provide quite a good text for a discourse on that interesting subject.

NOTES AND REFERENCES

For the analysis of variance where subclass frequencies are unequal, see Brandt (1933) and an important paper by Yates (1934*a*). Wilks (1938*e*) has considered the subject from the theoretical viewpoint and exhibited the main results determinantally. For the missing plot technique see Allan and Wishart (1930) and Yates (1933*b*). For the analysis of covariance see Fisher's *Statistical Methods*, Bartlett (1934*a*), an appendix by E. S. Pearson to a paper by Wilsdon (1934), Brady (1935), Wishart (1936), and Day and Fisher (1937). The last-mentioned paper works through a practical example in some detail and will repay study.

See also references to the previous chapter.

EXERCISES

24.1. For a two-way classification with one member in each subclass show that, for normal variation,

$$E\,(x_{j.} - x_{..})\,(x_{.k} - x_{..}) = 0,$$

and hence that the sums $\sum_j (x_{j.} - x_{..})^2$ and $\sum_k (x_{.k} - x_{..})^2$ are independent. Examine how this breaks down for the non-orthogonal case.

24.2. Verify the arithmetic of Example 24.6.

24.3. Generalise formula (24.73) in the following way. If there are m independent variates, the variance of corrected differences is

$$s^2 \left\{ \frac{2}{q} + \sum_{r,\, s=1}^{m} c_{rs}\, \lambda_r\, \lambda_s \right\}$$

where $\lambda_r = x_{rj} - x_{rk}$, and $c_{rs} = \dfrac{A_{rs}}{A}$ where A_{rs} is the cofactor of a_{rs} in the determinant $|\, a_{rs} \,|$, and $a_{rs} = \Sigma\, x_r\, x_s$ summed over the sample.

(Wishart, 1936.)

24.4. Derive by the analysis of variance the test of a regression coefficient given in **22.19**.

CHAPTER 25

THE DESIGN OF SAMPLING INQUIRIES

Influence of Theory on Sampling Design

25.1. The reader who is accustomed to handling the results of a sampling investigation as they appear in everyday statistical work may have wondered more than once in previous chapters whether theory was not reaching out too far in advance of practice. It is true that for certain types of experimental inquiry, notably in agricultural and biological research, the precision of exact statistical tests does not seem out of place; but in economic or social statistics, for example, there is often so much error and imperfection in the raw data that the application of refined methods of analysis would be a waste of time. It is clearly useless, and may even be dangerous, to exercise an elaborate mathematical technique on data which are suspect from the very start of the inquiry. If our theory is to be really serviceable to the statistician and not merely an enticing mental exercise it must be capable of solving practical problems.

25.2. Now it has to be admitted that much of the material with which statisticians have to work at the present day cannot be treated by the methods expounded in the foregoing pages when sampling questions are concerned. The commonest reason, but by no means the only one, is that the sampling process by which the data were obtained was biassed. In such cases the statistician has to lay aside the refined implements of his craft and do the best he can with his refractory material in the light of his own judgment and commonsense. A good deal of current statistical work is of this kind, and there is even a section of thought which is inclined to depreciate the advanced theory of the subject as "academic" in the sense that it is too remote from practical affairs to be worth studying. The misunderstanding is not likely to be removed by the counter-accusation sometimes launched by theoreticians that the theory is quite capable of being applied by anyone who has the ability to comprehend it.

25.3. Fortunately there is a growing realisation that the two points of view can often be reconciled by collecting the data in such a form that the theory *can* be applied to it. If only enough care is taken at the initial stages of an inquiry there is no need for the appearance of imperfect data which defy exact analysis. Knowing beforehand what theoretical instruments are at our disposal, and armed with a clear understanding of what questions we are trying to answer, we can frequently frame the investigation so as to maximise the information acquired with the minimum of effort. In short, the scope and nature of our theory itself dictates, to some extent, the form which the sampling inquiry should assume. In former times the statistician was usually asked to extract information from data which were collected by inexpert agents, frequently for quite different purposes. Nowadays he is still in the same position in some respects, but sometimes he is called in to advise on the design of the inquiry and can, within limits, determine the form in which the data are collected. He can make his theory applicable by selecting his sample in the proper way.

25.4. The general theory of the design of sampling inquiries has not progressed far enough for us to be able to give a systematic account of it in this chapter. In some fields,

particularly that of agricultural experimentation, it has reached quite an advanced degree of perfection ; in others there remain many problems unsolved and possibly many more which have not yet even been formulated. At the risk of some discontinuity of treatment, therefore, we shall only give in this chapter a number of instances in which theoretical considerations exert a considerable effect on the scope of a sampling inquiry, in order to illustrate the field to be covered. There are, of course, many factors which ultimately determine the form of an investigation, such as cost and expenditure of time, but they will not concern us here. For the present we shall be concerned solely with the extent to which theoretical considerations contribute to all the factors that have to be taken into account when an inquiry is designed.

Some Preliminary Points

25.5. There are certain preliminary points which, though obvious enough when stated explicitly, are often overlooked and cause a good deal of bad design.

(*a*) The fundamental object of sampling is to obtain information about a population, and it is of the first importance to begin with a clear idea of what that population is. Imagine, for instance, that we are asked to ascertain whether pasteurised milk has a different feeding value from raw milk. In what population is this inquiry to be made : among children ? among the inhabitants of the British Isles ? among those who habitually drink milk or those who do not ? among townspeople or among country folk ? and so on. Again, suppose that we are given a new variety of barley and wish to know whether it has a heavier yield than a previously known type. Do we mean heavier in the usual barley-growing areas ? in every kind of climate or on the average over a series of different climatic conditions ? when subject to the same manurial treatments as those in current use ? and so on.

(*b*) In a similar way, it is necessary to have an equally clear idea of what we are trying to find out about the population. In our example of raw and pasteurised milk, are we content to know that there is (or is not) a differential effect for children as a whole ? or do we wish to ascertain whether any such effect varies at different ages, between sexes, or according to nutritional standards ? What exactly should we like to know ? It is no use returning the facile reply " all about it " to this query, for our information must be limited in virtue of the finite size of our sample. We must make up our minds what information we require and which questions have priority if it becomes necessary to sacrifice some of them for practical reasons.

(*c*) Thirdly, we should consider what we know already about our population. This point becomes of particular importance when our prior knowledge indicates heterogeneity, for then we may, in effect, have to divide the population into sub-groups and sample separately from each. In our milk example, it is to be expected that children of different ages may react differently, or that children from lower-class schools may respond differently from those in middle-class schools. Or again, in our barley example, the two varieties may compare quite differently on Hertfordshire loam and on Lincolnshire chalk. It would be misleading to lump all the comparisons together when we have strong reason to suspect heterogeneity beforehand. In effect, prior knowledge of this kind frequently dictates the types of question we ask under (*b*), and the two are often different facets of the same problem.

(*d*) As an extension of the same point, we may notice that prior knowledge about the population sometimes indicates what sort of averages to use and what sort of tests of significance it is proper to apply. Crop-yields, for instance, are known to be distributed in a form approaching the normal, so that arithmetic means are good estimates of parent

means and the tests based on normal theory may be applied. Accident statistics, on the other hand, are often distributed in a modified Poisson form ; income statistics in a J-shaped form, and so forth.

(*e*) A specification of the population and a decision as to the precise object of the inquiry will usually determine certain parameters which it is required to estimate or certain hypotheses for test. In general the problem is one of estimation, but not necessarily so. In our case of pasteurised and raw milk, for instance, we should probably wish to know the exact amount of the difference between the effects of the two (a matter of estimation), not merely whether a difference existed (a matter of significance). We then wish to know, before the inquiry begins, whether the estimates we shall have are going to be accurate enough for our purpose ; or alternatively, if the sample is of a given size, how accurate they will be. It may not always be possible to answer such a question completely beforehand, since the sampling variances will in general depend on quantities which have to be estimated when the data are available, but it is always useful to consider in a general way what sort of magnitudes would be shown as significant and what values would leave us still in reasonable doubt. As a rule, matters such as this are closely related to sample size.

(*f*) Finally, our estimates will be subject to experimental error and, in development of the last point, we have to try to find the form of experimental design which, while answering our questions, does so with the minimum error. From a slightly different standpoint, if we can determine the amount of error which is admissible, the problem is to find the design which achieves no more than that error with the minimum expenditure of effort. Furthermore, we require to be able to estimate the extent of probable errors. In short, we require an *efficient* design, just as the engineer requires an efficient engine or the aircraft designer an efficient form of airscrew, and for exactly the same reasons.

25.6. To sum up, our primary task in embarking on a sampling inquiry is to ascertain as accurately as possible what is the population under examination, and what is the information about it which we require. If, as usually is the case, that information concerns statistical characteristics such as means and variances, or more generally frequency-distributions, our second task is to design an inquiry which will provide estimates of these unknown quantities and will, at the same time, provide estimates of their sampling error. It is not always possible, as we shall see later, to obtain full satisfaction in the reduction of error and the estimation of error simultaneously. Increased accuracy of estimation may mean loss of precision in our estimate of sampling error, so that although we are nearer the truth we do not know how near. There does not appear to be any single rule which will cover all the cases that can arise. We shall refer to a particular case of some interest in **25.39.**

Stratified Sampling

25.7. We consider at the outset a case of fairly frequent occurrence in the sampling of existent populations. Suppose we are interested in the mean value of a variate x in some population Π ; and that we know, or suspect, that the population is heterogeneous in the sense that we can delimit sub-populations $\Pi_1, \Pi_2, \ldots \Pi_k$ in which the distributions according to x may differ. This type of case might, for example, arise if we were sampling the population of a town for income, there being districts, wards or even streets which are known to be inhabited by classes living at different income-levels.

Practical considerations alone may require that we draw a prescribed portion of the sample from each sub-population. For instance, with a town of 500,000 inhabitants it

would be most tedious to sample by using random numbers applied to the whole town. We should probably divide the work among districts and blocks and select random samples within the blocks. This, however, is not to be confused with the division of the town into relatively homogeneous districts because of its heterogeneity. Either process is called *stratification*. The problem we shall discuss is this : If we have decided to draw a total sample of n members, and can assign at will the number n_i drawn from the ith stratum Π_i, subject to the condition $\Sigma\,(n_i) = n$, how should we choose the numbers n_i, or need we choose them at all ? Will our estimate of the mean value of x be better if we merely choose n members at random from Π, or can we improve it by controlling the numbers n_i and not merely leaving them to chance ?

25.8. Let x_{ij} be the jth member of the sample from the ith sub-population, and let the latter contain a number N_i of members with mean μ_i and variance σ_i^2. If μ is the mean of Π we shall have

$$\mu = \frac{1}{N}\sum_{i=1}^{k} N_i\,\mu_i. \qquad (25.1)$$

We shall now seek for parameters λ_{ij} such that our estimator of μ, say t, is given by

$$t = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (\lambda_{ij}\,x_{ij}), \qquad (25.2)$$

that is to say, is a linear estimator in the observed variate-values. We shall seek for that estimator which is unbiassed and has minimum variance, i.e. for which

$$E\,(t) = \mu \qquad (25.3)$$

$$E\,\{t - E\,(t)\,\}^2 = \text{minimum}. \qquad (25.4)$$

Substituting from (25.2) and (25.1) in (25.3), we find

$$E\left\{\sum_{i,j}\lambda_{ij}\,x_{ij}\right\} = \frac{1}{N}\sum_i N_i\,\mu_i$$

and since $E\,(x_{ij}) = \mu_i$ this gives

$$\sum_i \mu_i\left(\sum_j \lambda_{ij} - \frac{N_i}{N}\right) = 0. \qquad (25.5)$$

For this to be generally true we must have

$$\sum_{j=1}^{n_i}\lambda_{ij} = \frac{N_i}{N}, \qquad (25.6)$$

a first condition on the λ's. If $\lambda_{i.}$ is the mean of λ_{ij} in the ith set we have

$$\lambda_{i.} = \frac{N_i}{Nn_i}. \qquad (25.7)$$

Now consider (25.4). The variance of t is the sum of k variances, for the samples from sub-populations are independent. Consider then the variance of $\sum_j \lambda_{ij}\,x_{ij}$, remembering

that the population of N_i members is finite. We have

$$\begin{aligned}
\text{variance} &= E \sum_j \{\lambda_{ij} (x_{ij} - \mu_i)\}^2 \\
&= \sum_j \lambda_{ij}^2 \sigma_i^2 + \sum_{j,k} \{E \lambda_{ij} \lambda_{ik} (x_{ij} - \mu_i)(x_{ik} - \mu_i)\}, \qquad j \neq k \\
&= \sum_j \lambda_{ij}^2 \sigma_i^2 - \sum_{j,k} \lambda_{ij} \lambda_{ik} \frac{\sigma_i^2}{N_i - 1} \\
&= \frac{\sigma_i^2 N_i \Sigma \lambda_{ij}^2}{N_i - 1} - \frac{\sigma_i^2 (\Sigma \lambda_{ij})^2}{N_i - 1} \\
&= \frac{\sigma_i^2}{N_i - 1} \{n_i (N_i - n_i) \lambda_{i.}^2 + N_i \Sigma (\lambda_{ij} - \lambda_{i.})^2\}. \qquad (25.8)
\end{aligned}$$

This is clearly minimised only if

$$\lambda_{ij} - \lambda_{i.} = 0, \qquad (25.9)$$

that is, if all the λ's for any sub-population are equal. This is what we should expect on intuitive grounds, for there is no reason for weighting the sample members differently in the same sub-sample.

Our minimal variance, say v, is then given from (25.8), by summing over i, as

$$\begin{aligned}
v &= \sum_i \frac{\sigma_i^2 (N_i - n_i)}{N_i - 1} n_i \lambda_{i.}^2 \\
&= \frac{1}{N^2} \sum_i \frac{\sigma_i^2 (N_i - n_i)}{N_i - 1} \frac{N_i^2}{n_i} \\
&= \frac{1}{N^2} \sum_i \frac{\sigma_i^2 N_i^3}{(N_i - 1) n_i} + \text{constant}. \qquad (25.10)
\end{aligned}$$

This is a minimum for variations in n_i subject to $\Sigma n_i = n$ if

$$\frac{\partial}{\partial n_i} (v - p \Sigma n_i) = 0,$$

where p is an undetermined constant. This yields almost at once

$$n_i^2 \propto \frac{\sigma_i^2 N_i^3}{N_i - 1}. \qquad (25.11)$$

25.9. If we know the population variances σ_i^2 and the numbers N_i this equation determines the numbers n_i; but in practice it is rather unlikely that we should know the variances without knowing the means, in which case we should not have to sample to find the mean of the whole population. Our result is not, however, useless. In the first place we find for the estimator t

$$\begin{aligned}
\sum_i \sum_j \lambda_{ij} x_{ij} &= \sum_{i,j} \frac{N_i}{N} \frac{x_{ij}}{n_j} \\
&= \sum_i \frac{N_i}{N} x_{i.} \qquad (25.12)
\end{aligned}$$

so that the estimate is a weighted average of the sample means, the weights being proportional to the population numbers N_i, not to the numbers n_i. Secondly, without knowing the variances σ_i^2 exactly, we may sometimes reach approximations from prior knowledge of the populations. Such values, without giving absolute accuracy, will at least represent improvements on selecting the n's by chance.

25.10. If the numbers N_i are effectively infinite the formulae simplify, and, for instance, instead of (25.11) we have

$$n_i \propto \sigma_i N_i, \qquad (25.13)$$

the sample number varying with the standard deviation in the stratum concerned, as well as its number of members.

25.11. If there is no information available at all about the variances σ_i^2 the most reasonable course in applying (25.11) appears to be to suppose them all equal. In such a case, for large N_i we have

$$n_i \propto N_i, \qquad (25.14)$$

or the sampling numbers are proportional to the population numbers. This is what we might expect on intuitive grounds. If the populations are infinite the n_i's are equal, which again is in accordance with intuitive ideas.

25.12. The above will serve as an illustration of the way in which theoretical requirements can influence the scope of an inquiry conducted among an existent population. By seeking for an estimator with minimum variance we have been led to expressions determining the allocation of sample numbers among the different strata—and incidentally, of course, we have derived expressions for the minimum variance, so that the maximum possible precision can be ascertained. The fact that some of our results depend on unknown constants suggests that in some circumstances it may be worth while conducting a preliminary or " pilot " inquiry in order to estimate the unknowns and hence to improve the precision of the main inquiry which is to follow. The possibilities of such pilot surveys have yet to be explored, but the technique appears to merit serious investigation.

25.13. In passing, we may mention one other topic of great practical importance on which theory can throw a good deal of light, that of optimum size of a sampling unit. In sampling a human population of a town, for instance, need we take individuals as our units ? It would be easier to sample households, or streets, or even whole districts ; but do we lose anything by this method, and if so, how much ? Furthermore, the grouping of individuals into units of larger size sometimes has a peculiar effect on correlations which may lead to erroneous conclusions, and a theoretical investigation may be required to safeguard against error. We shall not pursue the subject further here—the sampling problem would require a book in itself—but the reader who is interested may like to consult some of the papers referred to at the end of the chapter.

The Design of Experiments

25.14. For an existent population the flexibility of sampling technique is somewhat limited. We are given an aggregate of values, some of which are to be extracted for scrutiny, and no manipulation of the sampling can tell us more than exists, so to speak, already inscribed upon the population itself. Consequently the main line of endeavour in such cases lies in estimating with the greatest accuracy (which is largely a matter of choosing the right statistics and minimising sampling variability), or in ensuring that sufficient material is available to enable the requisite comparisons to be made with significance (which is largely a matter of sample size and selecting the most suitable tests of significance). Nothing can alter the population, and theory will, as a rule, only react upon the sampling process by some such method as has already been exemplified, e.g. in dictating that the

sampling must be random, in stratifying the population before the sampling is carried out, and in deciding how limited resources can be expended to the best advantage.

25.15. For hypothetical populations there are often wider possibilities, for the nature of the inquiry may itself determine which populations are to be studied, and the populations may, to a certain extent, be set up at will. For instance, if we are interested in an inquiry into the relationship between income and size of family in the United Kingdom, the population already exists and we cannot go outside it ; whereas if we wish to discuss the effect of a poison on bacterial growth or of a fertiliser on the yield of barley we can not only reproduce experimental data *ad libitum* but can arrange the inquiry so as to confine it to certain populations (e.g., by considering only a given type of bacterium in fixed nutritional circumstances or at fixed temperatures), or we may extend the domain of consideration as far as purely practical limitations will allow (e.g., by growing barley in new surroundings or in new climates). This is rather a pretentious way of saying that we may *experiment* in a domain which, within limits, can be assigned at will. The statistician has a much greater scope for ingenuity in the design of experiments than in the design of sampling inquiries on existent populations because of the greater degree of control over the population under examination.

25.16. In the classical ideal experiment, only the factors under consideration were allowed to vary, other conditions being kept as constant as laboratory practice would allow —in investigations concerning the relation between resistance and current in an electric circuit, for instance, attempts would be made to keep factors such as temperature and external magnetic effects strictly constant. It would be recognized that there would be residual errors which would affect the exactitude of the results, but these would be measurable on certain assumptions.

25.17. Statistical theory can, of course, deal with such cases, but it can also go farther and often wishes to do so. In the first place, it frankly admits the existence not only of experimental error (in the sense of aberration from a " true " value) but of the much wider type of variation which gives rise to frequency-distributions in practice. Instead of isolating particular factors for study, it may wish to give full play to the disturbances which arise in practice in order to investigate what happens in " natural " conditions. For this reason, statistical experiments are often complex in the sense that a number of factors are allowed to vary simultaneously.

Secondly, the admission of outside influences which together make up what is generally called experimental error implies that it should be possible to estimate the extent of such error from the data themselves. We wish to obtain, not the functional relations between variables which may only exist under artificial conditions, but the stochastic relations observed in practice.

25.18. The effect of this on experimental design is that the hypothetical population we consider is often a rather general one. Taking the case of trials of a new variety of barley as an example, we should wish to compare its yields with those of other varieties in different soil conditions, with different manurial treatments, in different years (so as to get variations in climate), and so on. Furthermore, to obtain estimates of the error due to other factors we usually have to replicate the experiment. A great number of intercomparisons fall to be made, and the process of design is essentially that of finding a form

of experiment which will permit all these comparisons and yet save as much unnecessary labour as possible.

Orthogonality

25.19. To reduce the discussion to more concrete terms we will consider the testing of a new variety of barley. In order to study its behaviour under different soil conditions we will select a number of areas in which barley is grown and choose a block of ground in each. This will give us inter-soil comparisons. We will also arrange to carry the experiment on for a period of years, so that climatic variations may also be compared. The other factor in which we are interested is the response to certain manures, which we will take to be dung (D), potash (K), nitrogen (N), and phosphates (P).

Consider any block at any one place in any year. We will decide on certain standard quantities of the four manures and assume that for any manure either a dressing of this standard amount is to be given, or it is to be withheld. This simplifies the experiment, for then every manure either is or is not applied, and our results can be classified by simple dichotomies. Of course more complicated experiments can be devised to allow for different quantities of fertiliser, but the simpler case will be sufficient for our purposes.

We have then set up a population which can be classified according to six qualities, place, time, and the application of four manures. Our results are intended to show whether there is any variation in yield between these conditions and various combinations of them. Of course, it does not follow in deductive logic that if there is significant variation from year to year in the particular years chosen there will always be temporal or climatic variation ; and similarly, if there is significant variation from place to place it does not follow that other soil conditions which have not been tested will show a significant variation. To arrive at such conclusions we have to perform an ordinary generalisation by induction. What we shall say, if significant results appear, is that in the regions tested, or for the years tested, there were significant variations, and that it therefore appears likely that soil and climate exert a material effect on yield—and we shall maintain this with more or less confidence according as our experience is wider or narrower. This is the familiar inductive inference which forms the basis of all scientific inquiry.

25.20. Within any one block we shall wish to study the effect of manurial treatments not only separately but in combination. We therefore divide the block into sixteen compartments and treat them, respectively, with no manure, D, K, N, P, DK, DN, DP, KN, KP, NP, KNP, DNP, DKP, DKN and $DKNP$. Here every possible combination appears once and only once. To compare, for instance, the mean yields in the presence or absence of dung we add all the eight yields for plots on which no dung was spread and compare it with the sum of the other eight. All the necessary comparisons can be made.

Data of this kind are said to be *orthogonal.* Each possibility arises an equal number of times. The reason for the use of the word is that such material is orthogonal in the sense we have considered in the analysis of variance. We saw in Chapters 23 and 24 that where cell-frequencies were equal the analysis was greatly simplified, and that under the customary hypotheses the estimates of means were independent. It is not, of course, absolutely necessary to have orthogonal data—in fact, we have shown in Chapter 24 how to deal with the non-orthogonal case ; but it is evidently a great convenience to be able to arrange for orthogonality, and no efficiency is lost by doing so.

Replication

25.21. If, as suggested above, we divide each block into 16 plots and treat each differently, the analysis of variance of any block will have 15 degrees of freedom; and if we cannot ignore any of the interactions there will be no residual variance due to " error ", that is to say we cannot estimate the reliability of our comparisons. All the 15 possible independent comparisons may be made, but we cannot decide whether differences are significant in the sense that they may be due to the other factors which we have agreed to allow to bear on the experiment, such as individual soil differences from plot to plot. If we are to estimate such " error " we must give the factors which produce it an opportunity of varying. This may be done by *replicating* the experiment, that is to say, by repeating it in the same form. For instance, suppose that we set up four blocks and divide each into 16 plots, applying our manurial treatments to each block. Then, assuming that there are no significant interactions between blocks and treatments (a matter which we can test by examining the interaction terms in the variance-analysis), we shall have 63 degrees of freedom, of which 15 are assignable to treatments and their interactions and the remaining 48 to a " residual " term, the latter providing an estimate of experimental error. We have exemplified this process in Chapter 23.

Randomisation

25.22. Up to this point we have said nothing about the arrangement of our 16 plots within the block. Suppose we divide our block into plots of equal size. Is there any advantage in allocating the treatments systematically, or is it preferable to assign them at random ?

We shall consider the relative merits of random and systematic arrangements in more detail below, but we can announce the general rule now : unless there is some good reason to the contrary, it is better to allot the treatments at random. Where possible, chance should be given full play.

25.23. The justification for this rule in our present instance can be seen by reference to the section on randomised blocks in **23.41**. We saw there that by randomising the allocation of plots we were able to preserve the z-distribution and hence to validate our tests of significance, even where normality in the parent form was not assumed. The process is essentially one of extending our hypothetical population. Instead of considering the observed yields as specimens of what might happen in repeated trials of the same variety of barley if the same manurial treatments were applied to the same plots, we consider the possible yields in repeated trials if the manurial treatments were applied in all possible ways to different plots. Our experiment is systematic in the sense that we prescribe a different treatment for each plot; it is random to the extent that we allot the treatments to plots by chance.

25.24. There is one source of possible confusion here which it is desirable to remove. In our agricultural example complications arise because of the physical contiguity of the plots, and we shall see below that it is often desirable to eliminate by special designs systematic fertility gradients in the soil. In other classes of experiment where we desire orthogonality, the members need not be subject to this kind of effect, and often are not. Reverting to the example of raw versus pasteurised milk which has already been mentioned, suppose we take a simplified case and wish to measure whether the two different milks have different

effects on boys and girls. With a class of 40 children, 20 boys and 20 girls, we can proceed in several ways. It is obviously useless to give raw milk to all the boys and pasteurised milk to all the girls, for then we have no measure of the differential effect, if any, for either sex alone. We might toss up in each case and allot raw or pasteurised milk to each child by chance; but this would probably make the data non-orthogonal. To attain orthogonality, we should allot 10 children to each of the four sub-groups BP, GP, BR, GR (where B = boy, G = girl, P = pasteurised, R = raw). We then have an analysis of variance—

	Degrees of freedom
Between sexes	1
Between milks	1
Residual (including interactions)	37
TOTAL	39

This is analogous to a test of a cereal with two fertilisers and 10 replications.

The question is, how should we allot the children to the four groups? Their sex, of course, is determined, but the nature of the milk they receive is at choice. It is here that the randomisation will help. The ten children of a specified sex who receive raw milk should be chosen at random from the 20 available. In this instance it might be thought that any method would do; but it is best to avoid the risk of bias. If the children were chosen by the teacher he might tend to select the 10 bigger boys or the 10 brighter boys. If they were chosen alphabetically, we might get brothers and sisters automatically receiving the same treatment; and so on. The randomisation process avoids all systematic effects of this kind and brings us a stage nearer to obtaining an unbiassed answer to our questions.

Sensitivity of a Test

25.25. In some cases, where the variate is discontinuous, the nature of the test of significance which we propose to apply may make a difference to the form of the experiment. If we are testing a certain hypothesis which can produce a specified number m of experimental results which are acceptable as conforming to the hypothesis, whereas other hypotheses produce a number n of other results, we clearly want to keep m as small as possible compared with n. The ideal case, of course, is that of the "crucial" experiment in which the hypothesis can only give one result and other hypotheses give a different result. The result then proves or disproves the truth of the hypothesis, and no test of significance arises. In statistical practice we do not as a general rule perform crucial experiments, but we can sometimes design an experiment so that it is more crucial, if the expression be allowed, than alternative methods.

25.26. Consider, for instance, the case of a cashier who claims to be able to detect good money from false at a glance. To test this ability we spread ten coins before him, tell him that p are good, and ask him to point them out. What number of good coins p should we include among the ten?

If the cashier had no power of discrimination and there are p good coins, the probability that he would guess right by chance is

$$1\Big/\binom{10}{p},$$

for the total number of ways of selecting p from 10 is the denominator of this fraction and only one of them is right. Now we want to choose p so as to minimise the probability of such an event, i.e. so as to maximise $\binom{10}{p}$. This is clearly done when $p = 5$, so that we ought to have five good and five bad coins in the set. Any other number would increase the probability that he might be right by chance and hence decrease the sensitivity of the experiment.

Latin Squares

25.27. We now proceed to consider a different type of design, which has been freely applied in agriculture but may also be applied to other forms of inquiry. Suppose we have a variety of barley to test and five different treatments to apply. We will assume that replication has been considered necessary and will replicate five times, the same number as the treatments. We will then divide our block into 25 plots like a chessboard (though the plots may be rectangular and need not be exact squares, provided they are all the same size). Each row may be considered a replication of the five treatments, and this itself involves the appearance of each treatment once and only once in each row. Can we extend the arrangement and ensure that in addition the treatments will occur just once in each column ?

The answer is affirmative, as the following example shows :—

$$\begin{matrix} A & B & C & D & E \\ B & C & A & E & D \\ C & E & D & A & B \\ D & A & E & B & C \\ E & D & B & C & A \end{matrix} \qquad . \quad . \quad . \quad . \quad . \quad (25.15)$$

An arrangement of this kind is called a " Latin square ". It was studied extensively by Euler in the eighteenth century, though not of course from the statistical viewpoint.

25.28. The advantage of this arrangement lies in the fact that it eliminates possible correlational effects due to fertility gradients in the soil or accidental circumstances which may exercise a " patchy " influence on the whole block. If we could be sure that there were no such influences at work, and that the soil was entirely homogeneous in the block, it would not matter where the treatments were placed ; but by imposing the restriction that no treatment appears more than once in the same row or column we remove at least horizontal and vertical gradients from our comparisons. Suppose in fact that there were gradients running across the block and down it. When we work out the mean yield of the treatment A we shall add together five values, one of each in the various rows and columns. Similarly for B, so that a comparison of A and B is not affected by the systematic influences, which work equally on both.

It is not, of course, true that the Latin square arrangement eliminates every effect due to soil heterogeneity. There might be systematic effects running diagonally which might still remain. It is, however, clear that in removing the effects in two perpendicular directions we have substantially improved the comparison of mean yields as compared with a systematic arrangement.

25.29. The analysis of variance of a $p \times p$ Latin square may be carried out in the following form :—

Sum of squares	*d.f.*
Between rows	$p - 1$
Between columns	$p - 1$
Between treatments	$p - 1$
Residual	$(p - 1)(p - 2)$
TOTAL	$p^2 - 1$

(25.16)

and the four constituent sums are, on the hypothesis of homogeneity, distributed as $v\chi^2$ independently. Before proving this result we will consider an example.

Example 25.1 (from Thomson, *Brit. J. Educ. Psych.*, 1941, **11**, 135 ; data by S. D. Nisbet).

A set of children were divided into four equal groups and each group was given four lists of words to test spelling ability. Each list formed one of four different types of test which we denote by A, B, C, D. The arrangement of the experiment is shown in the following table, together with the total scores of the corresponding groups :—

Groups of children

		1	2	3	4	TOTALS
Lists of words	1	A 81	B 41	C 44	D 53	219
	2	D 38	A 97	B 42	C 49	226
	3	C 31	D 43	A 67	B 36	177
	4	B 57	C 33	D 43	A 81	214
	TOTALS	207	214	196	219	836

For instance, the first group of children had the first list of test A, the second of test D, and so on. No group had the same lists as another group, and each list was used exactly once. The scores (corresponding to yields in the agricultural case) were in fact the number of words spelled wrongly in a prior test but correctly in this test.

The above table, of course, does not represent anything corresponding to the physical layout of an agricultural experiment, but it shows how a similar object can be secured to the avoidance of contiguous effects. Since it is possible that some relationship may exist between the lists of words and the tests (e.g. by accident one list might be particularly unsuitable for a test), we wish to ensure that not only will each group of children have each of the four tests, but that no list shall be given more than once and every list at least once. This is precisely what the Latin square accomplishes. The fact that the diagonal arrangement of the letters is systematic does not affect the present inquiry, though in an

agricultural experiment a systematic diagonal fertility gradient might affect comparisons between treatments.

An analysis of variance on the usual lines gives the following results :—

Sum of Squares.		d.f.	Quotient.
Lists (rows)	359·5	3	119·83
Groups (columns)	74·5	3	24·83
Tests (treatments)	4626·5	3	1542·17
Residual	606·5	6	101·08
TOTALS	5667·0	15	

The differences between lists are evidently not-significant, from which we should conclude that they appear to be on a par so far as these tests are concerned. The quotient due to groups indicates that the children are more alike than chance would lead us to expect, but not significantly so, for the variance ratio $101{\cdot}08/24{\cdot}83 = 4{\cdot}1$, $\nu_1 = 6$, $\nu_2 = 3$, is not significant. On the other hand, the quotient due to tests is very significant, the ratio $1542{\cdot}17/101{\cdot}08 = 15{\cdot}3$, $\nu_1 = 3$, $\nu_2 = 6$ being beyond the 1-per-cent. point. We conclude that there do exist differences between the tests.

Construction of Latin Squares

25.30. The numbers of possible Latin squares of order p is very large for high values of p. There are, for example, 576 squares of order 4 ; 161,280 squares of order 5 ; 373,248,000 of order 6 and 61,428,210,278,400 of order 7. Up to this order they have been enumerated. Although many examples of squares of higher orders are known, the problem of enumeration for $p \geqslant 8$ awaits solution. Details and examples will be found in Fisher and Yates' *Statistical Tables*.

By interchanging rows and columns the square can always be brought to a form in which the top row and left-hand column are in the order ABC, etc. It is then said to be a " standard square ". For instance, there are four standard squares of the fourth order :—

$$\begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & B & A \\ D & C & A & B \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ B & C & D & A \\ C & D & A & B \\ D & A & B & C \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ B & D & A & C \\ C & A & D & B \\ D & C & B & A \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{array} \quad . \qquad (25.17)$$

From each of these, 144 ($= 4\,!\;3\,!$) squares may be derived by permuting all columns and all rows except the first. (There is no point in permuting the first row, because the result would be a repetition of squares already obtained with an interchange of the letters A . . D, not an essentially different layout.) The total number of squares, as stated above, is therefore $4 \times 144 = 576$.

It is only necessary to specify the standard squares. To select a Latin square at random we choose a standard form at random and then permute rows and columns at random, the randomising process being most conveniently carried out by Sampling Numbers. For squares of order 8 or more, where the standard types have not been enumerated, we can only choose one of those which has, and hence select one at random from a restricted set of all possible squares.

Analysis of Variance for Latin Squares

25.31. We must now justify our assertion that the Latin square may be analysed in the form (25.16), and that the z-test applies to the variance ratios which arise in the analysis.

For an ordinary two-way classification we have

$$\Sigma\,(x_{jk} - x_{..})^2 = \Sigma\,(x_{j.} - x_{..})^2 + \Sigma\,(x_{.k} - x_{..})^2 + \Sigma\,(x_{jk} - x_{j.} - x_{.k} + x_{..})^2.$$

Thus, if x_r is the mean of rows and x_c that of columns in the Latin square, we have, writing $\bar{x}$ for $x_{..}$,

$$\Sigma\,(x_{rc} - \bar{x})^2 = \Sigma\,(x_r - \bar{x})^2 + \Sigma\,(x_c - \bar{x})^2 + \Sigma\,(x_{rc} - x_r - x_c + \bar{x})^2 \quad . \quad (25.18)$$

and the three parts on the right are distributed independently as $v\chi^2$ with $p - 1$, $p - 1$ and $(p - 1)(p - 1)$ degrees of freedom respectively.

Now

$$\Sigma\,(x_{rc} - x_r - x_c + \bar{x})^2 = \Sigma\,(x_t - \bar{x})^2 + \Sigma\,(x_{rc} - x_r - x_c - x_t + 2\bar{x})^2 + 2\Sigma\,(x_t - \bar{x})\,(x_{rc} - x_r - x_c - x_t + 2\bar{x}) \quad . \quad . \quad (25.19)$$

where x_t is the mean of treatments.

Consider the cross-product term in (25.19). The summation takes place over all p^2 values in the Latin square. Let us confine our attention to the summation for some particular treatment. For this summation the factor $x_t - \bar{x}$ is constant. Summation for the other factor gives

$$\Sigma\,(x_{rc} - x_r - x_c - x_t + 2\bar{x}) = px_t - \Sigma\,x_r - \Sigma\,x_c - px_t + 2p\bar{x} \quad . \quad (25.20)$$

and since one treatment occurs in each row and column,

$$\left.\begin{aligned}\Sigma\,x_r &= p\bar{x}\\ \Sigma\,x_c &= p\bar{x},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25.21)$$

and hence the sum (25.20) vanishes.

Thus the cross-product in (25.19) vanishes also and we have

$$\Sigma\,(x_{rc} - \bar{x})^2 = \Sigma\,(x_r - \bar{x})^2 + \Sigma\,(x_c - \bar{x})^2 + \Sigma\,(x_t - \bar{x})^2 + \Sigma\,(x_{rc} - x_r - x_c - x_t + 2\bar{x})^2. \quad . \quad . \quad (25.22)$$

This gives us the analysis of the sums of squares, and it only remains to show that the third term on the right in (25.22) is independent of the fourth. It will then follow that the four terms are distributed independently with $p - 1$, $p - 1$, $p - 1$ and $(p - 1)(p - 2)$ degrees of freedom.

The required property of independence can be established directly, but it also follows from considerations of symmetry in the Latin square which have an interest of their own. We have regarded the square as composed of rows and columns, with treatments allotted in a certain way; but by rearrangement we can equally well regard it as composed of rows and treatments with columns allocated in a certain way. For instance, if we take the first standard square in (25.17) we may write it :—

	Treatment : A	B	C	D
Rows : 1	C_1	C_2	C_3	C_4
2	C_2	C_1	C_4	C_3
3	C_4	C_3	C_1	C_2
4	C_3	C_4	C_2	C_1

where, for instance, treatment A occurs in row 1, column 1 (C_1), row 2, column 2 (C_2), and

so on. This, of course, is not a physical layout, but that is immaterial for present purposes. It follows that since the sum of squares between columns is independent of the residual in (25.22), so also is that between treatments.

The variance analysis then takes the form

Sum of Squares.		d.f.
Rows	$\Sigma (x_r - \bar{x})^2$	$p - 1$
Columns	$\Sigma (x_c - \bar{x})^2$	$p - 1$
Treatments	$\Sigma (x_t - \bar{x})^2$	$p - 1$
Residual	$\Sigma (x_{rc} - x_r - x_c - x_t + 2\bar{x})^2$	$(p - 1)(p - 2)$
TOTALS	$\Sigma (x_{rc} - \bar{x})^2$	$p^2 - 1$

. (25.23)

25.32. The above form provides a homogeneity test of the usual kind. If the test proves significant of heterogeneity we may, in the usual way, consider the hypothesis that

$$x_{rc} = a_r + b_c + c_t + \zeta_{rc} \qquad (25.24)$$

where ζ_{rc} is normally distributed about zero mean. We leave it to the reader to show, as in Chapter 23, that in such an event the residual mean square is an unbiassed estimate of the variance of ζ with $(p - 1)(p - 2)$ degrees of freedom.

25.33. As in the case of randomised blocks, it appears that under certain general conditions the z-distribution is reproduced approximately for fixed values which are permuted in all the permissible ways consistent with the Latin square design. We omit an investigation into this result (for which see Welch, 1937) as the algebra is considerably more complicated than for randomised blocks. The result has been confirmed by a limited number of experiments.

Graeco-Latin and Orthogonal Squares.

25.34. If the two squares

$$\begin{array}{cccc} A & B & C & D \\ B & A & D & C \\ C & D & A & B \\ D & C & B & A \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ C & D & A & B \\ D & C & B & A \\ B & A & D & C \end{array} \qquad (25.25)$$

are superposed we have the arrangement—

$$\begin{array}{cccc} AA & BB & CC & DD \\ BC & AD & DA & CB \\ CD & DC & AB & BA \\ DB & CA & BD & AC \end{array} \qquad (25.26)$$

in which every possible pair of letters (XY being regarded as different from YX) appears just once. Such a pair of squares is said to be orthogonal. The form (25.26) is sometimes written with Greek letters instead of the second Roman set; hence the name of Graeco-Latin square. It is also possible to superpose a third factor which we will denote by the

numerals 1–4 in such a way that each combination of any pair of types occurs just once, e.g.

$$\begin{array}{cccc} A\,\alpha\,1 & B\,\beta\,2 & C\,\gamma\,3 & D\,\delta\,4 \\ B\,\gamma\,4 & A\,\delta\,3 & D\,\alpha\,2 & C\,\beta\,1 \\ C\,\delta\,2 & D\,\gamma\,1 & A\,\beta\,4 & B\,\alpha\,3 \\ D\,\beta\,3 & C\,\alpha\,4 & B\,\delta\,1 & A\,\gamma\,2 \end{array} \quad . \quad . \quad . \quad . \quad (25.27)$$

Complete sets of orthogonal squares (i.e. those in which there are $p - 1$ factors for a $p \times p$ square) are known for all prime p and for $p = 4$, 8 and 9. Curiously, there is no set for $p = 6$. Up to and including $p = 7$ they have been enumerated.

We shall not enter here into the use of these squares in experimental design. They are generalisations of the Latin square in which, by suitable arrangements, several factors can be tried out simultaneously, so that all possible combinations of pairs occur an equal number of times.

Confounding

25.35. It will be evident that if we wish to consider in full a classification according to several variates, particularly with replications, the number of individual members in the sample may be very large. For instance, if we wish to test a variety of barley with three different applications of four types of fertiliser, there must be 81 yields even without replication, if we want to make all the comparisons possible. Physical considerations may make a layout of an experiment on such a scale impossible. The difficulty is possibly more serious in experiments on expensive animals such as cows.

Where economy in the size of sample is a very material factor we may be able to reduce the sample at the expense of sacrificing some of the less important comparisons. For example, to consider once again the case of barley and the effect of fertilisers: we shall undoubtedly wish to compare yields of D and not-D, K and not-K, P and not-P, N and not-N. We may also wish to compare first-order interactions of the type DK and not-D, K. But it is quite possible that interactions of higher order, such as the effect of dung in the presence of two other fertilisers, are negligible. Where we are prepared to assume that this is so, on the basis of prior evidence or otherwise, we can dispense with certain information and still make the comparisons we wish while retaining properties of orthogonality.

25.36. Consider, as an illustration, an experiment with three fertilisers, each of which is applied or not applied, say N, P and K, and four replications. In the ordinary way there would be 32 plots and we should have an analysis of variance as follows, assuming that block-treatment interactions may be regarded as part of the residual :—

Sum of squares.	d.f.
Blocks	3
N	1
P	1
K	1
NP	1
NK	1
PK	1
NPK	1
Residual	21
TOTAL	31

Now suppose that we divide our main blocks into two sub-blocks, the first containing the treatments

$$O \text{ (None)}, NP, NK, PK, \quad . \quad . \quad . \quad . \quad . \quad (25.28)$$

and the second the treatments

$$N, P, K, NPK. \quad . \quad . \quad . \quad . \quad . \quad (25.29)$$

We may then analyse the variance as follows, regarding the sub-blocks as blocks of four plots each :—

Sum of squares	d.f.
Blocks	7
N	1
P	1
K	1
NP	1
NK	1
PK	1
Residual	18
TOTAL	31

In fact, if we wish to compare the yields with N and those without N, i.e.

$$N + NPK + NP + NK$$

with

$$O + PK \quad + P \quad + K,$$

it will be seen that we add two members from (25.28) and two from (25.29), so the difference is not affected by block differences ; and similarly for the other comparisons. Such a design is said to be *balanced*, and the interaction NKP is *confounded* with block-differences, since in the eight blocks it cannot now be isolated from block effects. The advantage of the second design over the first is that, without losing anything appreciable in comparisons between treatments, we have gained a good deal in the assessment of block effects ; for the residual has only declined from 21 to 18 d.f. whereas the sum of squares between blocks has increased from 3 to 7 d.f.

25.37. The ideas of orthogonality, randomisation, balance and confounding have been developed to an advanced degree and with great ingenuity, particularly by Fisher and Yates. The slight sketch we have given of the methods in this chapter is intended to be no more than illustrative of the way in which the theory of experimental design is capable of development, at least in certain fields, and the manner in which efficiency may be imported into a practical inquiry by a due regard to theoretical requirements of the design. For a comprehensive account of this branch of the subject the reader should consult Fisher's *Statistical Methods* and *Design of Experiments*, Yates (1937*b*), and a useful introductory account by Goulden (1939). At this point we leave these particular topics and return to certain general matters.

Design and Randomisation

25.38. Whenever an inference is to be made, and particularly where hypothetical populations are concerned, the reader will find it useful to ask himself what precisely is the population under consideration. We can illustrate the point very usefully by discussing

a subject on which there has recently been difference of authoritative opinion—that of occasional conflict between the requirements of balancing and randomisation.

25.39. Consider in the first place the testing of a cereal under two treatments, denoted by A and B; and to simplify matters as much as possible, suppose we are to sow eight plots in a straight line. In what order shall we allot the treatments?

If the plots are not too large so that the row covers a big area, it is quite possible that there may be a trend of fertility in the soil itself which will affect yields differentially and hence interfere with comparisons which we might make. Suppose that we do wish to guard against a fertility gradient so far as possible. We might then decide on one of the "balanced" arrangements:

$$A\,A\,B\,B\,B\,B\,A\,A \qquad . \quad . \quad . \quad . \quad . \quad (25.30)$$

$$A\,B\,B\,A\,A\,B\,B\,A \qquad . \quad . \quad . \quad . \quad . \quad (25.31)$$

$$A\,B\,A\,B\,B\,A\,B\,A \qquad . \quad . \quad . \quad . \quad . \quad (25.32)$$

As will be easily seen, if there is a linear gradient in fertility along the row the means of A and B treatments respectively will be affected to the same extent and hence their difference unaffected. For instance, consider (25.30) and suppose the linear gradient is represented by an additive factor $q + kp$, $k = 1 \ldots 8$. On the hypothesis that the remaining effect consists of a constant a for A-treatments with a normal residual ξ, and similarly for B, the yields are

A-treatments: $q + p + a + \xi_1,\ q + 2p + a + \xi_2,\ q + 7p + a + \xi_7,\ q + 8p + a + \xi_8$
B-treatments: $q + 3p + b + \xi_3,\ q + 4p + b + \xi_4,\ q + 5p + b + \xi_5,\ q + 6p + b + \xi_6$

with means

$$\tfrac{1}{4}(4q + 18p) + a + \tfrac{1}{4}(\xi_1 + \xi_2 + \xi_7 + \xi_8)$$
$$\tfrac{1}{4}(4q + 18p) + b + \tfrac{1}{4}(\xi_3 + \xi_4 + \xi_5 + \xi_6)$$

respectively. The differences of these two are independent of q and p.

25.40. The alternative procedure in allotting treatments would be to distribute them at random. Such balanced arrangements as (25.30)–(25.32) might then arise by chance. But we might also get such an arrangement as

$$A\,A\,A\,A\,B\,B\,B\,B \qquad . \quad . \quad . \quad . \quad . \quad (25.33)$$

What are we to do in such circumstances? If we reject this arrangement we are rejecting the random allocation of treatments in favour of systematisation. If we accept it we know quite well that a fertility gradient, if it exists, will invalidate the inquiry.

The reader will no doubt agree that, if other things are equal, the balanced arrangement is better than the arrangement (25.33). What we have to examine is whether other things *are* equal; in short, whether in rejecting randomisation we have lost anything useful in the testing of significance.

25.41. Consider a rather more general case in which an experimental area is laid out in p blocks of q treatments each. If the subscript j refers to blocks and k to treatments, we have the usual analysis with sum of squares between blocks ($p - 1$ d.f.), between treatments ($q - 1$ d.f.), and residual ($(p - 1)(q - 1)$ d.f.).

Now we have seen that if the individual plot-yield can be regarded as a block effect plus a treatment effect plus a normal residual with constant variance from plot to plot,

the significance of treatment effects can be judged from the z-test in the usual way by comparing sum of squares between treatments with the residual sum of squares. This is true whether treatments are allocated at random or not.

But suppose we wish to adopt the alternative viewpoint of **23.41** and make the inference in the set of values obtained by permuting the observed values. These permutations will not affect the block means or the total mean, and hence the sum of squares between blocks remains constant. The remaining part of the analysis may be written—

	Sum of Squares.	d.f.
Treatment . . .	$S_1 = \Sigma (x_{.k} - x_{..})^2$	$q - 1$
Residual	$S_2 = \Sigma (x_{jk} - x_{j.} - x_{.k} + x_{..})^2$	$(p - 1)(q - 1)$
TOTALS . .	$S_3 = \Sigma (x_{jk} - x_{j.})^2$	$p (q - 1)$

. (25.34)

Rather remarkably, the z-test holds for the ratio

$$\frac{S_1}{q-1} \frac{(p-1)(q-1)}{S_2},$$

provided that treatments are allocated at random, independently of the distribution of residual effects in individual plots.

25.42. Consider, then, the population of values, $(q\,!)^{p-1}$ in number, obtained by permuting the observed values. The total sum of squares S_3 in (25.34) is the same for all members. Consequently if S_1 is too great, S_2 must be too small and vice-versa ; and in general, if we confine ourselves to certain layouts and reject others, all the possible values of S_1 cannot appear. It is this fact which has been seized on by advocates of randomisation. They point out that for balanced layouts S_1 tends to be smaller than for random layouts (a conclusion supported by experiment) ; consequently that the test of significance is invalidated and the estimate of error S_2 too big. The difference between the two modes of thought may be expressed briefly in this way : with balanced layouts the real error is reduced but the estimate of error is too large, so that the significance of the result is more in doubt ; whereas with random layouts the estimate of error is exact but the error itself may be larger. The question is whether one prefers to be nearer the truth without knowing how near, or farther from the truth with a knowledge of the limits of error.

25.43. For details of the controversy on this topic the reader may consult the papers referred to at the end of the chapter. It brings into prominence an important question of inference which can only be decided by the experimenter himself. If he chooses to regard any act of experimentation as one of a large population of such acts, to be carried out by himself or other workers, he may prefer randomisation in all circumstances, notwithstanding that every now and again he will hit by chance on a design which he knows is likely to give misleading results. But if he cannot take this very detached attitude (and most experimenters, being human, would think it poor compensation that their own errors are balanced by the better luck of other people) then he will prefer to design a balanced layout, even if the exactitude of his tests of significance is impaired.

25.44. We must, however, not leave the reader with the impression that the desiderata of both schools of thought are totally incompatible. It frequently happens that one can select a design which is both balanced and random. The Latin square is a good example. By imposing the restriction that a treatment must not appear more than once in a row or column we remove to some extent the interference of fertility gradients; by requiring that it shall appear just once we balance the design; and by leaving the rest of the layout to be determined by a random selection from all possible Latin squares of that order we randomise so as to reproduce the distribution of the variance ratio in the required form, thus, as "Student" remarked, "conforming to all the principles of allowed witchcraft".

REFERENCES

A classical case of how an inquiry can be spoilt by poor design is the Lanarkshire Milk Investigation, for which see "Student" (1931*c*) and E. M. Elderton (1933). This case will repay study. On some theoretical problems arising from the sampling of existent populations see Bowley (1925), Jensen (1925), Sukhatme (1935), Neyman (1933*b*, 1934, 1938*a*, 1939*a*, 1941*b*), Olds (1939, 1940), and Frankel and Stock (1939). The war has accentuated many of the points remaining unsolved, and there is much of general interest in recent issues of the *Journal of the American Statistical Association* and the *Annals of Mathematical Statistics*. For some work on the "pilot" sampling technique see Sukhatme (1935) and C. Bose (1943).

Reference has been made in the text to Fisher's *Design of Experiments*, Yates' *Principles of Orthogonality and Confounding*, and Goulden's *Methods of Statistical Analysis*.

For the problem of size of sampling units see the papers by Neyman referred to above, particularly 1934, and for its effect on correlation analysis see an interesting appendix in Wold's *Analysis of Stationary Time Series*.

For the controversy on balance versus randomisation see "Student" (1938), Barbacki and Fisher (1936), E. S. Pearson (1937*b*, 1938), and Jeffreys (1939*e*).

EXERCISES

25.1. A population is given by specifying the frequencies in comparatively narrow ranges of one variate, the frequency in the ith range being N_i and ranges being of equal width. Show that if the population frequencies are large, the best estimator of the mean of a second variate which is linearly related to the first (in the sense of the unbiassed estimator of minimum variance) in a sample obtained by taking n_i members from the ith range is given when n_i is proportional to N_i.

25.2. Extend the result of the previous exercise to the case where ranges are of unequal width.

If the number of farms in England and Wales is known in the acreage ranges 0–49, 50–99, 100–199, 200–499, 500 and over, what sampling proportions would you take in the various ranges to estimate the total acreage under wheat?

25.3. If a variate ξ can be regarded as the sum of a systematic component $\xi(x)$ and an uncorrelated random component ε_1 and η similarly as $\eta(x) + \varepsilon_2$, and if the random components are uncorrelated with each other, show that

$$r(\xi, \eta) = \frac{\operatorname{cov}\{\xi(x), \eta(x)\}}{\{(\operatorname{var} \xi(x) + \operatorname{var} \varepsilon_1)(\operatorname{var} \eta(x) + \operatorname{var} \varepsilon_2)\}^{\frac{1}{2}}}.$$

Hence, if a population is divided into strata the correlation between ξ and η for these strata will, in general, be less than that obtained by combining strata to obtain larger units; and as the strata are further subdivided the correlation between ξ and η tends to zero.

(Spearman, 1907, *Am. J. Psych.*, **18**; Wold, 1938*a*.)

25.4. Illustrate the effect of the foregoing exercise by calculating the correlation coefficients for the data of Table 14.4 (vol. I, p. 333), (*a*) by adding the variates in pairs and so obtaining 24 values; (*b*) by repeating the operation and obtaining 12 values; and (*c*) by repeating the operation and so obtaining 6 values.

25.5. (Markoff's theorem.) Consider a sample of n independent values $x_1 \ldots x_n$, x_i being drawn from a population Π_i with mean μ_i and variance σ_i^2. Suppose we have a function θ defined by

$$\theta = \sum_{j=1}^{s} b_j p_j$$

where the b's are known and the parameters p_j depend on the μ's according to the equation

$$\mu_i = \sum_{j=1}^{s} a_{ij} p_j, \qquad s \leqslant n$$

the a's also being known. Then an unbiassed estimator of θ, say t, with minimum variance may be written—

$$t = \sum_{j=1}^{n} \lambda_j x_j.$$

Show that the function t is given by substituting for the p's in the expression for θ the functions q given by minimising

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\left\{x_i - \sum_{j=1}^{s} (a_{ij} q_j)\right\}^2$$

with regard to the q's considered as independent variables.

Show further that if this minimum value is S_0 the estimated variance of t is

$$\frac{S_0}{n-s} \Sigma (\lambda_i^2 \sigma_i^2).$$

25.6. In a feeding experiment there are given five different foods, each of which is available in four grades. It is desired to feed each animal with one grade of each food, but only one, so that a comparison may be made of the effect of the different grades of any particular food. Use the Graeco-Latin square to show how the feeding can be carried out.

25.7. A water diviner is to be taken to ten spots and asked to say whether water is present below the surface. It is decided to choose five spots where water is known for certain to exist and five where it is known not to exist. The order in which the spots are to be presented is determined by spinning a coin, heads denoting water and tails not-water.

The spinning of the coin results in the first five trials giving heads. Would you accept this result or spin again ?

25.8. Show that a Latin square may be regarded as a three-way classification in which p^2 members are not zero, but $p^3 - p^2$ members vanish. Derive the analysis of variance for the Latin square from this approach and generalise it to the Graeco-Latin square.

CHAPTER 26

GENERAL THEORY OF SIGNIFICANCE-TESTS—(1)

Hypotheses to be Considered

26.1. The kind of hypothesis which we test in statistics is more restricted than the general scientific hypothesis. It is a scientific hypothesis that every particle of matter in the universe attracts every other particle, or that Homer was blind; but these are not hypotheses such as arise for testing from the statistical viewpoint. A review of the various tests which have been introduced earlier in this book indicates that the great majority specify something about a population. Some merely assert a general fact such as "the population is continuous" or "the population is rectangular". Others are more definite, as for instance "the population is normal and has a mean μ_0"; and again others are less definite in one direction and more definite in another, e.g. "the population has unit variance". It is also usually a part of the hypothesis that the sample from which the inference is being made was obtained by a random process.

26.2. Suppose we have a set of random variables $x_1 \ldots x_n$. In the sample space W of n dimensions the sample-point whose co-ordinates are $x_1 \ldots x_n$ determines a point E, say, with a distribution function which we may write as $P(E)$. If w is any region in W, we may derive the probability that E falls in w, say $P(E \varepsilon w)$. Then we shall say that any hypothesis concerning the law $P(E \varepsilon w)$ is a *statistical hypothesis*. If it determines the law completely we shall call it *simple*. In the contrary case it is said to be *composite*.

For instance, in testing the significance of the mean of a sample of n, it is a statistical hypothesis that the parent is normal. This is composite, as also is the hypothesis that the parent is normal with mean μ or the hypothesis that the parent is normal with variance σ^2. The hypothesis that the parent is normal with mean μ and variance σ^2 is simple because then the parent is fully determined.

Example 26.1

In sampling from a population dichotomised into classes possessing the attributes A or not-A, say in proportion ϖ and χ $(= 1 - \varpi)$, the sampling distribution is the binomial $(\chi + \varpi)^n$. This is completely determined by the value of ϖ, and hence a hypothesis as to the value of ϖ is simple. Such, for instance, would be the hypothesis that male and female births occur in equal proportions. Similarly, in a multiple classification with proportions $\varpi_1, \varpi_2, \ldots \varpi_s$, a simple hypothesis would specify values for all the ϖ's; if only one were specified and s were greater than two the hypothesis would be composite.

In sampling from a bivariate normal population characterised by two means, two variances and a correlation, a hypothesis about any one parameter would be composite, and similarly for a hypothesis concerning two, three or four parameters. Only if all five were specified in addition to the normality of the parent would the hypothesis be simple; and this notwithstanding the fact that the sampling distribution of the means is independent of the other three parameters, and that of the correlation coefficient independent of the other four.

26.3. A hypothesis which determines the law $P(E \varepsilon w)$ completely except for ν parameters is sometimes said to have ν degrees of freedom. Such a hypothesis may be regarded as an aggregate of simple hypotheses. For instance, the hypothesis that a population is normal with mean μ is the aggregate, for all σ^2, of hypotheses that it is normal with mean μ and variance σ^2.

26.4. The kind of argument we have used in testing hypotheses, for both large and small samples, is of this character: assuming that the hypothesis is true, we can, with any assigned probability α, find a region w_0 in the sample space W such that the probability of E falling in $W-w_0$ is α. We call $W-w_0$ the *region of acceptance* and the complementary domain w_0 the *critical region.* (This is the nomenclature of Chapter 19.) If our observed E falls in w_0 we reject the hypothesis; if not we accept it. As a rule, in practical cases, our regions w_0 are determined by the values of some statistic such as $\bar{x}$ in testing the mean.

Errors of First and Second Kind

26.5. In general, as we saw in Chapter 19, there are many possible regions of acceptance for any given hypothesis and any given probability level α. For all of them we shall err in proportion $1 - \alpha$ of the cases in the long run by rejecting the hypothesis if E falls in the critical region—*provided that the hypothesis is true.* But what about the case when it is not true? We cannot ignore this case, for its possible existence is the very reason for carrying out the test. It is of no use whatever to know merely what the test will do when the hypothesis is true without regard to its behaviour in the contrary case; for if we are to consider only the events which happen when the hypothesis is true we have no right to use a test based on that assumption to reject it.

By having regard to the behaviour of the test when the hypothesis is not true we are able to lay down criteria for choosing among the various tests obeying the rule

$$P\{E \varepsilon w_0 \mid H_0\} = 1 - \alpha, \qquad (26.1)$$

where H_0 is the hypothesis. In fact we shall seek for the test which, while obeying (26.1), minimises the risk of accepting H_0 when an alternative hypothesis H_1 is true and H_0 accordingly is false. That is to say, we shall endeavour to find w_0 such that, in addition to (26.1), we also have

$$1 - P\{E \varepsilon w_0 \mid H_1\} = \text{minimum}. \qquad (26.2)$$

26.6. From a slightly different viewpoint we may say that there are two possible errors in judging a statistical hypothesis:

(*a*) We may reject it when we ought to accept it, that is, when it is true.

(*b*) We may accept it when we ought to reject it, that is, when it is false.

These are known as errors of the first and second kind respectively. The error of the first kind we can control exactly by setting up the proper region of acceptance determined by α. Errors of the second kind cannot be controlled in this way, but we can sometimes calculate their probabilities, and in any case can try to reduce them to a minimum. This is the fundamental idea, first given explicit expression by Neyman and E. S. Pearson, which determines most of the work in the present and succeeding chapters.

26.7. The possibility of finding regions of acceptance obeying (26.2) clearly depends on a precise specification of what alternative hypotheses are under consideration. We had better emphasise the importance of this point. It is customary to speak, and even,

in a loose kind of way, to think of testing a hypothesis without reference to alternatives. To take the case of testing for normality, we often say that the hypothesis under test is that the population is normal without specifying what other form it might have. The reader may say that the alternative he has in mind is merely the negation of the hypothesis, namely that the population is not normal. But if so he will find it very difficult—in my own view impossible—to justify any of his tests on a logical basis. He will calculate certain statistics and accept the hypothesis if their values are consonant with the normal values; but it will always be possible to find other populations for which the observed values are even closer to expectation. If agreement between theoretical and observed values is the criterion he should reject normality in favour of these alternative hypotheses. It is not until he specifies his alternatives and considers errors of the second kind that some firm foundation for intuitive processes begins to appear.

26.8. Perhaps it may help to clarify the fundamental concepts of the present approach

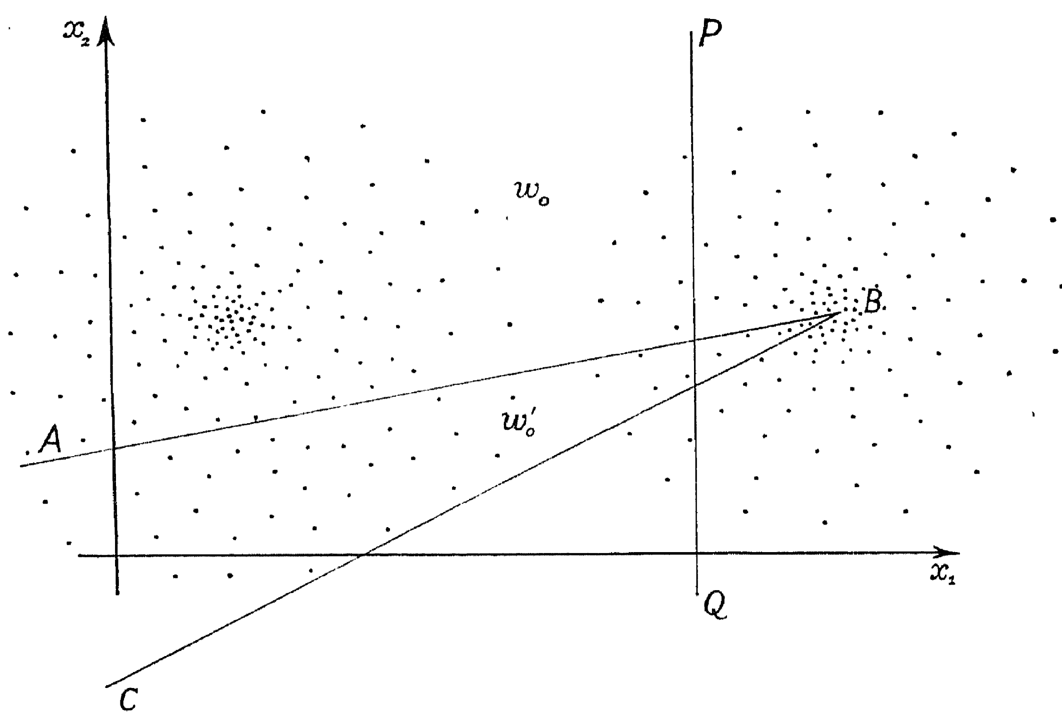


FIG. 26.1 (see text).

if we consider a simple illustration where the hypothesis under test H_0 is simple and there is only one alternative H_1 which is also simple. In Fig. 26.1 we show diagrammatically the scatter of sample-points which would arise in samples of two, x_1 and x_2, the cluster on the right being that due to H_0 and the one on the left to H_1. In practice, of course, the sampling distributions are more usually continuous, but the dots will indicate roughly the condensation of sample density round central values.

In determining the critical region we have to find an area in the (x_1, x_2) plane such that its "content" is $1-\alpha$. Two possible areas are shown, w_0 being the area to the left of the line PQ, and w_0' the area between the lines AB and BC. In either case the proportion in the critical regions of the frequency on hypothesis H_0 is $1-\alpha$, and if we reject H_0 whenever the sample-point falls in w_0 (and similarly for w_0') we shall commit an error of the first kind in proportion $1-\alpha$ of the cases in the long run.

Consider errors of the second kind. By using the region w_0 we should reject H_0—and

therefore accept H_1—every time the sample-point arose from H_1, that is to say in practically all the cases where H_1 was true, since nearly all the sample-points arising from H_1 lie in w_0. Errors of the second kind are therefore very rare. On the other hand, if we were to use w_0' we should accept H_0 every time a sample-point arose from H_1 but did not fall between the lines AB and BC, that is to say fairly frequently. Clearly w_0 is the better critical region and has a much smaller error of the second kind than w_0'.

26.9. It is to be noted that the argument does not depend on the relative frequencies of occurrence of the hypotheses H_0 and H_1. This is generally true. There is no concealed form of Bayes' postulate in this approach.

26.10. When there are n variates and p unknown parameters the geometrical representation can be extended by imagining a sample-space W of n dimensions adjoined to a parameter space of p dimensions. We cannot draw a picture of such a case on a two-dimensional sheet of paper, but the geometrical imagery and terminology of the method are frequently useful. A graphical illustration of a two-dimensional sample-space and a one-dimensional parameter space has already been given in Fig. 19.3.

The Power Function

26.11. If for a simple hypothesis H_0, (26.1) is true we define

$$P\{E \varepsilon w_0 \mid H_1\} = \beta(H_1 \mid w_0) \quad . \quad . \quad . \quad . \quad (26.3)$$

as the *power* of the critical region w_0 with respect to H_1. Clearly the power is greatest when the probability of an error of the second kind is least.

In the expression on the left of (26.3) we regard the probability that E falls in w_0 as dependent on H_1, the hypothesis alternative to H_0. In the expression on the right we have regard to the power of the test for H_1 as dependent on w_0.

If there exists a particular region w_0 with greater power than any other region obeying (26.1) we shall say that it is the best critical region, and the test based on it will be called the most powerful test.

26.12. We proceed to consider in turn the following cases :—

(*a*) H_0 simple ; one alternative H_1 which is simple.

(*b*) H_0 simple ; an alternative H_1 which is composite but can be regarded as an aggregate of simple alternatives.

(*c*) H_0 and H_1 composite but expressible as aggregates of simple hypotheses.

Simple Hypotheses : One Simple Alternative

26.13. Suppose the parent population is continuous, so that the simultaneous distribution of the n sample values $x_1 \ldots x_n$ is continuous ; and let the frequency functions of the sample values on hypotheses H_0 and H_1 be $p_0(x_1 \ldots x_n)$ and $p_1(x_1 \ldots x_n)$ respectively. Write dx for the element $dx_1 \ldots dx_n$. Then we have

$$\int_{w_0} p_0 \, dx = 1 - \alpha \quad . \quad . \quad . \quad . \quad . \quad (26.4)$$

and wish to maximise, for variations in the domain w_0, the integral

$$\int_{w_0} p_1 \, dx. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.5)$$

This is a problem in the Calculus of Variations and is equivalent to maximising unconditionally the integral

$$\int_{w_0}\left(p_1 - \frac{1}{k}p_0\right) dx, \qquad \qquad (26.6)$$

or, what is the same thing, to minimising

$$\int_{w_0}(p_0 - kp_1)\, dx, \qquad \qquad (26.7)$$

where k is a constant to be determined by (26.4).

It is known that the condition for a stationary value of (26.7) is that, on the boundary of w_0,

$$p_0 - kp_1 = 0. \qquad \qquad (26.8)$$

If the solution is a minimum we have, inside w_0,

$$p_0 \leqslant kp_1 \qquad \qquad (26.9)$$

and outside w_0,

$$p_0 > kp_1. \qquad \qquad (26.10)$$

This solution to the problem is fairly obvious on general grounds. If U is a function which is sometimes positive and sometimes negative, with a line of demarcation where it is zero (as must exist in virtue of continuity), we clearly minimise $\int U\, dx$ by taking into the region w_0 all the points for which U is negative and no more. This gives us (26.9) and (26.10), and the boundary of w_0 is the locus for which U vanishes. By convention we regard the boundary as included in w_0, which accounts for the equality in (26.9) and its absence in (26.10).

26.14. The conditions expressed by (26.8), (26.9) and (26.10) are sufficient as well as necessary. For let w_1 be any other region for which

$$\int_{w_1} p_0\, dx = 1 - \alpha.$$

If w_0 and w_1 have a common part denote it by w_{01}. Then

$$\begin{aligned}\int_{w_0 - w_{01}} p_0\, dx &= 1 - \alpha - \int_{w_{01}} p_0\, dx \\ &= \int_{w_1 - w_{01}} p_0\, dx\end{aligned}$$

and hence, from (26.9)

$$\begin{aligned}k\int_{w_0 - w_{01}} p_1\, dx \geqslant \int_{w_0 - w_{01}} p_0\, dx &= \int_{w_1 - w_{01}} p_0\, dx \\ &> k\int_{w_1 - w_{01}} p_1\, dx.\end{aligned}$$

Adding to both sides $k\int_{w_{01}} p_1\, dx$, we have

$$k\int_{w_0} p_1\, dx > k\int_{w_1} p_1\, dx, \qquad \qquad (26.11)$$

and hence, for positive k, the power of w_1 is less than that of w_0 and the latter is the best critical region.

Both in this section and implicitly in the last we have required k to be positive. That it must be so if w_0 is to exist emerges from (26.8), for p_0 and p_1 are essentially not negative, and if k were negative no solution for real variate-values would exist.

Example 26.2

Consider the normal population

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp\{-\tfrac{1}{2}(x-\mu)^2\}\,dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Let the hypothesis H_0 be that $\mu = a_0$, and the alternative that $\mu = a_1$. We have—

$$p_0 = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{n}(x_j - a_0)^2\right\}.$$

We can conveniently express this in terms of the sample mean $\bar{x}$ and the sample variance s^2, obtaining for the density function

$$p_0 = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left[-\frac{n}{2}\{(\bar{x}-a_0)^2 + s^2\}\right].$$

A similar expression is found for p_1 and thus, for the boundaries of the best critical region, we have

$$\frac{1}{k} = \frac{p_1}{p_0} = \exp\left[-\frac{n}{2}\{(\bar{x}-a_1)^2 - (\bar{x}-a_0)^2\}\right]$$

$$= \exp\left[-\frac{n}{2}(a_0 - a_1)(2\bar{x} - a_0 - a_1)\right].$$

This yields for the critical region

$$(a_0 - a_1)(2\bar{x} - a_0 - a_1) \leqslant \frac{2}{n}\log k,$$

or

$$(a_0 - a_1)\bar{x} \leqslant \tfrac{1}{2}(a_0^2 - a_1^2) + \frac{1}{n}\log k = (a_0 - a_1)\bar{x}_0, \text{ say.}$$

If $a_1 < a_0$ the region is then defined by

$$\bar{x} \leqslant \bar{x}_0,$$

but if $a_1 > a_0$ it is defined by

$$\bar{x} \geqslant \bar{x}_0.$$

The reader should compare the two cases on a diagram similar to that of Fig. 26.1.

Example 26.3

Consider again the normal population when the mean is known, say zero, but the variance unknown, e.g.—

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{x^2}{2\sigma^2}\right)dx, \qquad -\infty \leqslant x \leqslant \infty.$$

We now find, for hypotheses $\sigma = \sigma_0$ and $\sigma = \sigma_1$

$$k = \frac{p_0}{p_1} = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left\{-\frac{n}{2}(\bar{x}^2 + s^2)\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right\}$$

which yields, for the best critical region,

$$(\bar{x}^2 + s^2)(\sigma_0^2 - \sigma_1^2) \leqslant \frac{2\sigma_1^2\sigma_2^2}{n} \log\left\{k\left(\frac{\sigma_0}{\sigma_1}\right)^n\right\}$$
$$\leqslant v\,(\sigma_0^2 - \sigma_1^2), \text{ say.}$$

Thus our critical regions are defined by

$$m_2' = \bar{x}^2 + s^2 \leqslant v \qquad \text{if} \quad \sigma_1 < \sigma_0$$
$$m_2' = \bar{x}^2 + s^2 \geqslant v \qquad \text{if} \quad \sigma_1 > \sigma_0$$

The best critical regions in the space W are thus bounded by hyperspheres centred at the origin. Whether we take the space inside or the space outside a particular hypersphere as the critical region depends on the alternative hypothesis. The probabilities concerned can be evaluated directly without evaluating the constants k and v. In fact, the probability of exceeding a given value of $\frac{nv}{\sigma_0^2} = \frac{n(\bar{x}^2 + s^2)}{\sigma_0^2} = \chi_0^2$ is obtainable from the χ^2-distribution with n degrees of freedom, and hence the relation between v and α can be ascertained from the χ^2-integral.

In this particular case we may find without difficulty the power of an alternative test which would suggest itself on intuitive grounds. Suppose we find $\frac{(n-1)v'}{\sigma_0^2} = \chi_1^2$ from the χ^2-distribution corresponding to $n - 1$ degrees of freedom and probability level α, and use, instead of the hyperspheres centred at the origin, those centred at the sample mean

$$s^2 \leqslant v', \quad s^2 \geqslant v'.$$

Suppose that the alternative H_1 is that $\sigma_1^2 = 1{\cdot}1\ \sigma_0^2$. In testing H_0 for the alternative $\sigma_1 > \sigma_0$ we should, for the test based on v, find χ_0^2 and accept σ_0 if

$$\frac{nm_2}{\sigma_0^2} \leqslant \chi_0^2.$$

For instance, with $n = 5$, $1 - \alpha = 0{\cdot}01$ we find $\chi_0^2 = 15{\cdot}086$. The probability of an error of the second kind is

$$\int_{w_0} p_1\, dx = \int_0^{\chi_0^2/1\cdot 1} dF\,(\chi^2),$$

i.e. is obtained from the χ^2-integral with argument $\frac{\chi_0^2}{1{\cdot}1} = 13{\cdot}71$, giving $\beta\,(H_1 \mid w_0) = 0{\cdot}018$.

On the other hand, had we used χ_1^2 instead of χ_0^2 we should have entered the table with four degrees of freedom, giving $13{\cdot}277$. Divided by $1{\cdot}1$ this gives $12{\cdot}07$, resulting in a probability of rather less than $0{\cdot}017$. This is the power of the second test and is lower than that of the first test, as of course it must be since the latter has maximum power.

Simple Hypotheses: Families of Simple Alternatives

26.15. Consider now the case where H_0 is simple but H_1 is composite and consists of a family of simple alternatives. The most frequently occurring case is the one in which we have a class of simple hypotheses Ω of which H_0 is one and H_1 comprises the remainder; for example, the hypothesis H_0 may be that a mean has some value μ_0 and the hypothesis H_1 that it has some other value unspecified.

For each of these other values we may apply the foregoing results and find for each α corresponding to any particular member of H_1, say H_t, a best critical region w_t. But this region in general will vary from one H_t to another. We obviously cannot determine a different region for all the unspecified possibilities and are therefore led to inquire whether there exists, among the family of best critical regions w_t, one which is the best for all of them. Such a region is called the Uniformly Most Powerful and the test based on it the Uniformly Most Powerful test, conveniently shortened to U.M.P. test.

26.16. Unfortunately, as we shall find below, the U.M.P. test does not usually exist unless we restrict our family Ω in certain ways. Consider, for instance, the case dealt with in Example 26.2. We found there that for $a_1 < a_0$ the best critical region for a simple alternative was defined by

$$\bar{x} \leqslant \bar{x}_0.$$

Now the boundaries of the regions determined by $\bar{x}$ = constant do not depend on a_1 and can be found directly from the sampling distribution of $\bar{x}$ when the probability level $1 - \alpha$ is given. Consequently the regions defined by $\bar{x} \leqslant \bar{x}_0$ are the same for all $a_1 < a_0$ and hence the test is U.M.P. for the class of hypotheses that $a_1 < a_0$. It is difficult to see how a better test could be devised, for, whatever a_1 subject to $a_1 < a_0$, the test controls errors of the first kind and minimises those of the second.

However, if $a_1 > a_0$ the best critical regions are defined by $\bar{x} \geqslant x_0$. Here again, if our class Ω is confined to the values of a_1 greater than a_0 the test is U.M.P. But if a_1 can be either greater or less than a_0 no U.M.P. test is possible. The reader will easily verify for himself that the same is true for the test considered in Example 26.3.

26.17. We now show formally that for a simple hypothesis depending on θ_0—the value taken by the parameter θ defining a family of alternatives—no U.M.P. test exists for both positive and negative values of $\theta - \theta_0$ if the frequency function $p\,(E \mid \theta)$ is continuous, has everywhere a continuous derivative with respect to θ which does not vanish identically, and admits of differentiation under the sign of integration over W.

Suppose that such a test does exist. Then for any θ we have, inside w_0

$$p_0 \leqslant kp,$$

which we may write

$$p\,(E \mid \theta) \geqslant h\,(\theta)\, p_0\,(E \mid \theta_0). \qquad (26.12)$$

Likewise, for any point $\bar{E}$ on the boundary of w_0 we have

$$p\,(\bar{E} \mid \theta) = h\,(\theta)\, p_0\,(\bar{E} \mid \theta_0). \qquad (26.13)$$

By hypothesis p is differentiable in θ and hence so is h. Moreover, as $\theta \to \theta_0$, $h\,(\theta) \to 1$. Hence if

$$\Delta = \theta - \theta_0$$

and primes denote differentiation with respect to θ, we have

$$\begin{aligned} h\,(\theta) &= 1 + \Delta\,[h']_{\theta_0+q\Delta} \qquad 0 \leqslant q \leqslant 1 \\ &= 1 + \Delta\left[\frac{\partial}{\partial\theta}\frac{p\,(\bar{E} \mid \theta)}{p_0\,(\bar{E} \mid \theta_0)}\right]_{\theta_0+q\Delta} \\ &= 1 + \frac{\Delta}{p_0\,(\bar{E} \mid \theta_0)}\,[p'\,(\bar{E} \mid \theta)]_{\theta_0+q\Delta} \end{aligned} \qquad (26.14)$$

Further we have

$$p(E \mid \theta) = p_0(E \mid \theta_0) + \Delta\,[p'(E \mid \theta)]_{\theta_0 + r\Delta} \qquad 0 \leqslant r \leqslant 1. \qquad . \quad (26.15)$$

Substituting in (26.12) from (26.14) and (26.15), we find

$$\Delta\left\{[p'(E \mid \theta)]_{\theta_0 + r\Delta} - \frac{p_0(E \mid \theta_0)}{p_0(\bar{E} \mid \theta_0)}[p'(\bar{E} \mid \theta)]_{\theta_0 + q\Delta}\right\} \geqslant 0 \quad . \quad . \quad (26.16)$$

This is true for any E and $\bar{E}$ and for all Δ, *whatever* its sign, and hence the expression in curly brackets vanishes. Thus we have

$$[p'(E \mid \theta)]_{\theta_0} - \frac{p_0(E \mid \theta_0)}{p_0(\bar{E} \mid \theta_0)}[p'(\bar{E} \mid \theta)]_{\theta_0} = 0. \quad . \quad . \quad . \quad (26.17)$$

Similarly this equation may be shown to hold outside w_0, and hence it is true throughout W.

Now we have

$$\int_W p(E \mid \theta)\, dx = 1,$$

and hence, differentiating with respect to θ and putting $\theta = \theta_0$,

$$\int_W [p'(E \mid \theta)]_{\theta_0}\, dx = 0.$$

Substituting from (26.17), we have

$$\int_W \frac{p_0(E \mid \theta_0)}{p_0(\bar{E} \mid \theta_0)}[p'(\bar{E} \mid \theta)]_{\theta_0}\, dx = 0,$$

and hence

$$\frac{[p'(\bar{E} \mid \theta)]_{\theta_0}}{p_0(\bar{E} \mid \theta_0)} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.18)$$

Thus, from (26.17)

$$[p'(E \mid \theta)]_{\theta_0} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.19)$$

But this implies that the derivative of p with respect to θ is identically zero at θ_0, which is contrary to hypothesis. The theorem follows.

It may be noted that in deriving (26.17) from (26.16) we used the property that Δ may have either sign. If it can have only one sign, that is, if our class of admissible alternatives is confined to the case when either $\theta < \theta_0$ or $\theta > \theta_0$, a U.M.P. test may exist; and so we found in Examples 26.2 and 26.3.

Best Critical Regions and Likelihood

26.18. Since on the boundary of a best critical region we have $p_0 - kp_1 = 0$, that boundary is determined by the condition that on it the ratio of the likelihoods of two functions corresponding to H_0 and H_1 is constant.

Consider now the case where H_1 comprises a set of alternatives varying according to the parameter θ, H_0 being one of them. In accordance with the principle of maximum likelihood we should obtain, as the most likely value of θ, the solution of

$$\left(\frac{\partial p}{\partial \theta}\right)_{\theta = \hat{\theta}} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.20)$$

where $\hat{\theta}$ is then expressed as a function of the variables. If this value is substituted in p, we obtain the distribution with greatest likelihood which may be written p (Ω max.). The surfaces of constant likelihood are defined for this distribution by

$$p_0 - \lambda p\,(\Omega \text{ max.}) = 0. \qquad . \quad . \quad . \quad . \quad (26.21)$$

Now these surfaces are, in fact, the envelopes of the family, varying with θ,

$$p_0 - kp_\theta = 0, \qquad . \quad . \quad . \quad . \quad . \quad (26.22)$$

for to obtain the envelope we differentiate with respect to θ, giving $\dfrac{\partial p}{\partial \theta} = 0$ and eliminate θ, leading back to (26.21). Thus, if there exists a best critical region (and hence a U.M.P. test) for all permissible alternatives H_θ, such a region will be the envelope with respect to such alternatives and will therefore be identical with a region defined by (26.21); and hence a test based on the principle of likelihood leads to best critical regions, if they exist.

If, as is more usual, there is no common best critical region, the ratio of the likelihood of H_0 to that of any particular H_θ is k. The surface (26.21) remains the envelope of the family of surfaces (26.22) for which $k = \lambda$.

Example 26.4

Consider once again the normal form, where both mean μ and variance σ^2 are specified and the admissible alternatives are that they can have any values, subject of course to the variance being positive. For any given μ_1 and σ_1 the best critical region will be given by—

$$\frac{p_0}{p_1} = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left[-\tfrac{1}{2}\left\{\Sigma\left(\frac{x-\mu_0}{\sigma_0}\right)^2 - \Sigma\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\}\right] \leqslant k$$

or

$$\Sigma\left(\frac{x-\mu_0}{\sigma_0}\right)^2 - \Sigma\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \geqslant 2\log\left\{k\left(\frac{\sigma_0}{\sigma_1}\right)^n\right\}.$$

This may be written in the form

$$n\,\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\sigma_0^2}\,\{(\bar{x}-\rho)^2 + s^2\} \geqslant \text{constant}$$

where

$$\rho = \frac{\mu_0\sigma_1^2 - \mu_1\sigma_0^2}{\sigma_1^2 - \sigma_0^2}.$$

Thus, if $\sigma_1 > \sigma_0$ we have

$$(\bar{x}-\rho)^2 + s^2 \geqslant v^2, \text{ say};$$

and if $\sigma_1 < \sigma_0$ we have

$$(\bar{x}-\rho)^2 + s^2 \leqslant v^2.$$

For any specified μ_1 and σ_1 the best critical regions are bounded by hyperspheres with radius $v\sqrt{n}$ and centre at $x_1 = x_2 = \ldots = x_n = \rho$. Owing to the fact that ρ varies with μ_1 and σ_1, there will not in general be a best common critical region and a U.M.P. test; and this remains true even if we limit our alternatives to $\sigma_1 < \sigma_0$ and $\mu_1 < \mu_0$ or by similar inequalities.

We may regard $\bar{x}$ and s as independent variables and represent the data on a two-way plane $(\bar{x}, s)$. The best critical regions are then seen to be bounded by circles with

centre $(\rho, 0)$ and radius v. Fig. 26.2 (adapted from Neyman and Pearson, 1933*c*) illustrates some of the contours for particular cases. A single curve, corresponding to a single probability level, is shown in each case.

Cases (1) *and* (2) : $\sigma_1 = \sigma_0$ and $\rho = \pm\infty$. The best critical region lies on the right of the line (1) if $\mu_1 > \mu_0$ and on the left of (2) if $\mu_1 < \mu_0$. This is the case discussed in Example 26.2.

Case (3) : $\sigma_1 < \sigma_0$, say $\sigma_1 = \frac{1}{2}\sigma_0$. Then $\rho = \mu_0 + \frac{4}{3}(\mu_1 - \mu_0)$ and the region lies inside the semicircle marked (3).

Case (4) : $\sigma_1 < \sigma_0$ and $\mu_1 = \mu_0$. The region is inside the semicircle (4).

Case (5) : $\sigma_1 > \sigma_0$ and $\mu_1 = \mu_0$. The region is outside the semicircle (5).

There is evidently no common best critical region for these cases. The regions of

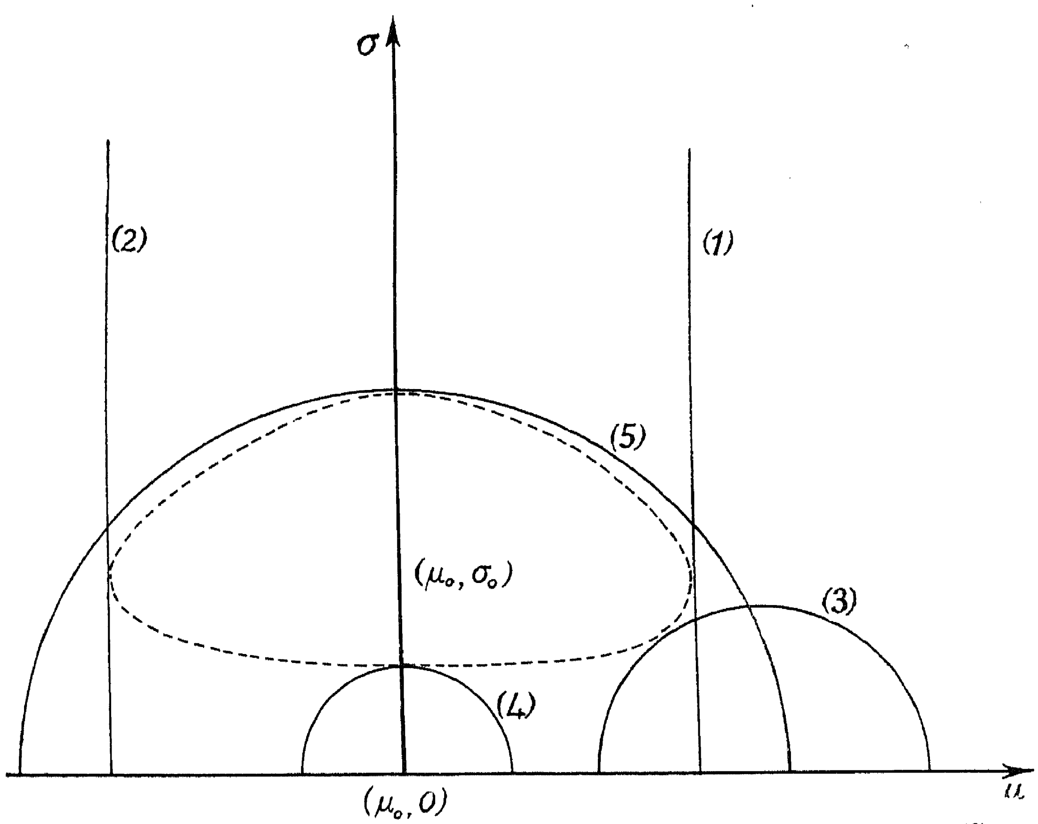


FIG. 26.2.—Contours of Constant Likelihood in a Two-dimensional Case. (See text.)

acceptance, however, may have a common part, centred round the value (μ_0, σ_0), and we should expect them to do so. Let us find the envelope of the best critical regions, which is, of course, the same as that of the regions of acceptance. The likelihood ratio is

$$k = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left[\frac{ns^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) - \frac{n}{2}\left\{\left(\frac{\bar{x} - \mu_0}{\sigma_0}\right)^2 - \left(\frac{\bar{x} - \mu_1}{\sigma_1}\right)^2\right\}\right].$$

The partial differentials with respect to μ_1 and σ_1 equated to zero give

$$\frac{n}{\sigma_1} - \frac{ns^2}{\sigma_1^3} - \frac{n}{\sigma_1}\left(\frac{\bar{x} - \mu_1}{\sigma_1}\right)^2 = 0$$

$$\frac{n}{\sigma_1^2}(\bar{x} - \mu_1) = 0,$$

whence we find $\mu_1 = \bar{x}$ and $\sigma_1 = s$ and the envelope is

$$1 - \frac{2}{n}\log k = \left(\frac{\bar{x} - \mu_0}{\sigma_0}\right)^2 - \log\left(\frac{s}{\sigma_0}\right)^2 + \frac{s^2}{\sigma_0^2}.$$

The dotted curve in Fig. 26.2 shows one such envelope. It touches the boundaries of all the critical regions which have the same likelihood-ratio k. The space inside may be regarded as a "good" region of acceptance and the space outside accordingly as a good critical region.

There is no best region for all alternatives, but the regions determined by envelopes of likelihood-ratio regions effect a sort of compromise by picking out and amalgamating parts of critical regions which are best for individual alternatives.

Example 26.5

In the previous example we have supposed that the sample space W was the same for all admissible alternatives. This is quite legitimate, for we can always regard the domain of variation as infinite by supposing that $p = 0$ outside the range of the frequency-distribution of the variates. In the normal case, of course, p does not vanish anywhere, so that we are compelled to consider W as infinite.

When, however, the sample-space for non-vanishing p is bounded, special circumstances may arise, and it is occasionally necessary to consider separately the different discriminating regions. For instance, if the sample-spaces corresponding to H_0 and H_1 are W_0 and W_1, it may happen that W_0 and W_1 have no common part when both p_0 and p_1 are greater than zero. If so, we can distinguish between H_0 and H_1 with certainty. If there is a common region W_{01} then $W_1 - W_{01}$ should be included in the best critical region, for to do so reduces the probability of errors of the first kind. But it does not follow that this should constitute the whole of the critical region, for we might then commit too many errors of the second kind, i.e. accept H_0 too often when H_1 is true. We may then wish to add to $W_1 - W_{01}$ a region w_{00}, making w_0 altogether, such that w_{00} lies inside W_{01} and $p_0\,(E\;\varepsilon\;w_{00}) = p_0\,(E\;\varepsilon\;w_0) = 1 - \alpha$. This controls the first kind of error to level α and reduces the second kind of error.

Consider the population

$$p\,(x) = \frac{1}{b}, \qquad a - \tfrac{1}{2}b \leqslant x \leqslant a + \tfrac{1}{2}b$$

$$= 0, \qquad \text{elsewhere.}$$

Suppose a sample of n to have been drawn from a population of this kind where b is known. We wish to test whether a has some value a_0 as against the alternative a_1.

The sample-spaces W_0 and W_1 are hypercubes centred at a_0 and a_1. If they have a common part W_{01} the probabilities p_0 and p_1 in that part are both proportional to the volume and $p_0/p_1 = 1$ everywhere in the region. If, then, we take any region w_{00} of content $1 - \alpha$ in W_{01} and add it to $W_1 - W_{01}$ we get a best critical region, and there are clearly infinitely many such.

For the admissible alternatives a_1 the hypercube W_1 will move along the long diagonal $x_1 = x_2 = \ldots = x_n$ as a_1 varies, and we cannot always find a common region of size $1 - \alpha$ to form w_{00}. By taking such a region as a hypercube of side $b\,(1 - \alpha)^{\frac{1}{n}}$, however, fitted into one of the corners of W_0 lying on the long diagonal, we "nearly" obtain such an object since this region provides what is required so long as W_0 and W_1 have a common part of content $1 - \alpha$. Which corner we choose depends on whether the hypothesis is $a_1 > a_0$ or $a_0 > a_1$.

Relation between U.M.P. Tests and Sufficient Estimators

26.19. It was thought at one time that the existence of a set of U.M.P. tests for a continuous range of admissible alternatives involved the existence of a sufficient estimator for the parameter concerned. This does not appear to be true in full generality, but is so in nearly all the cases occurring in statistical practice. We will prove a theorem on the subject :—

If a system of U.M.P. tests exists and if any point in the sample-space lies on the boundary of a best critical region, then a sufficient estimator exists for the parameter whose variation provides the admissible alternatives.*

It is enough to show that for an arbitrary point we have

$$p_1(E) = h(t, \theta)\, p_0(E) \qquad . \quad . \quad . \quad . \quad . \quad (26.23)$$

for then t is sufficient for θ by definition. Now we know that on the boundary of a critical region we have

$$\frac{p_1(E)}{p_0(E)} = \frac{1}{k} = h, \text{ say,}$$

where h varies with the x's and with θ. We show that h has the form $h(t, \theta)$ by defining a function t and showing that if t has the same value at any two points E_1 and E_2, then

$$\frac{p_1(E_1)}{p_0(E_1)} = \frac{p_1(E_2)}{p_0(E_2)}$$

for all θ.

26.20. For this purpose we require a lemma to the following effect : if a set of U.M.P. tests exists, it will be said to be ordered if the condition $\alpha_1 > \alpha_2$ implies that the critical region $w(\alpha_1)$ is included in the region $w(\alpha_2)$; and if a set of U.M.P. tests exists but is not ordered we can always find another set which is.

$w(\alpha_1)$ and $w(\alpha_2)$ may include parts of W where p vanishes. Let the remaining parts be $v(\alpha_1)$ and $v(\alpha_2)$ and, if v_0 is the common part of these regions, write

$$\left.\begin{aligned} v(\alpha_1) &= v_0 + v' \\ v(\alpha_2) &= v_0 + v'' \end{aligned}\right\} \qquad . \quad . \quad . \quad . \quad . \quad (26.24)$$

where v_0, v' and v'' have no common points. Now for any value of θ and for any E in $w(\alpha_1)$ —and therefore in v'—there is an h_1 such that

$$\begin{aligned} p_1(E) &\geqslant h_1\, p_0(E) \text{ in } v' \\ &\leqslant h_1\, p_0(E) \text{ outside, and therefore in } v''. \end{aligned}$$

Similarly, within $w(\alpha_2)$ and hence within v'' we have an h_2 such that

$$\begin{aligned} p_1(\dot{E}) &\geqslant h_2\, p_0(E) \text{ in } v'' \\ &\leqslant h_2\, p_0(E) \text{ in } v'. \end{aligned}$$

It follows that, from the inequalities deriving from v'', $h_1 \geqslant h_2$, and similarly, from v', $h_2 \geqslant h_1$. Hence $h_1 = h_2 = h$, say, and

$$p_1(E) = h\, p_0(E) \qquad . \quad . \quad . \quad . \quad . \quad (26.25)$$

within v' and v'' for any θ.

* The theorem remains true if there is a set of points of measure zero for which the condition as to boundaries is not fulfilled. It is also true for several parameters, as may be seen by an easy generalisation of the argument. See Neyman and Pearson (1936*a*).

Now take

$$u(\alpha_1) = v_0 + v''' \qquad . \quad . \quad . \quad . \quad . \quad (26.26)$$

such that

$$\int_{u(\alpha_1)} p_0 \, dx = 1 - \alpha_1. \qquad . \quad . \quad . \quad . \quad . \quad (26.27)$$

This is always possible, for the integral of p_0 over $v_0 + v''$ is $1 - \alpha_2$, which is greater than $1 - \alpha_1$. It follows from (26.27) and the first equation of (26.24) that

$$\int_{v'''} p_0 \, dx = \int_{v'} p_0 \, dx. \qquad . \quad . \quad . \quad . \quad . \quad (26.28)$$

Now put

$$w'(\alpha_1) = W_0 + u(\alpha_1) = W_0 + v_0 + v''',$$

where W_0 is the part of W for which $p_0 = 0$. Then from (26.27)

$$\int_{w'(\alpha_1)} p_0 \, dx = 1 - \alpha_1.$$

Further, $w'(\alpha_1)$ is a best critical region with respect to admissible alternatives, for (26.25) and (26.28) imply that

$$\int_{v'''} p_1 \, dx = \int_{v'} p_1 \, dx,$$

and hence

$$\int_{w'(\alpha_1)} p_1 \, dx = \int_{v(\alpha_1)} p_1 \, dx.$$

Finally, $w'(\alpha_1)$ is wholly included in $w(\alpha_2)$.

We have therefore replaced the region $w(\alpha_1)$ by another region $w'(\alpha_1)$ with the same properties except that it is included in $w(\alpha_2)$. The lemma follows.

26.21. To return now to the main proposition, let E be any point of W. If it belongs to only one boundary of a best critical region with content $1 - \alpha$ we put $t(E) = 1 - \alpha$. If it belongs to more than one, we put $t(E)$ equal to the mean between the upper and lower bounds of values of $1 - \alpha$ for which the boundaries include E. In virtue of the lemma, this implies that whatever the value of $1 - \alpha$ between these bounds, the corresponding boundary must contain E.

Thus t is defined everywhere. Further, if it has the same value at two points E_1 and E_2 these points must lie on the same boundary. It follows that on this boundary

$$\frac{p_1(E_1)}{p_0(E_1)} = \frac{p_1(E_2)}{p_0(E_2)}$$

and hence the theorem is proved.

The converse is not generally true, but one has to exercise some ingenuity and import some artificiality to construct examples where it fails. Cf. Exercises 26.3 and 26.4.

Composite Hypotheses

26.22. We shall consider a class Ω of admissible hypotheses depending on $r + s$ parameters $\theta_1 \ldots \theta_r \ldots \theta_{r+s}$ and shall regard the hypothesis H_0 under test as one of this class. A composite hypothesis of r degrees of freedom is one for which s of the parameters, say $\theta_{r+1} \ldots \theta_{r+s}$, are specified, the hypotheses determining the distribution apart from the unspecified parameters. For example, the hypothesis that a population

is normal with specified mean, nothing being supposed about the variance, is a composite hypothesis of one degree of freedom. It will be assumed that any admissible simple alternative is given by specifying the other r parameters $\theta_1 \ldots \theta_r$ and that there is a common sample-space W for all such alternatives.

Regions Similar to the Sample Space

26.23. In order to test the composite hypothesis H_0 we need in the first place to control errors of the first kind by determining a critical region w, such that

$$\int_w p_0 \, dx = 1 - \alpha. \qquad (26.29)$$

This, however, differs from the simple case in that p_0 can vary according to the unknown parameters, and to be certain of controlling the error we must be able to find w such that (26.29) is true whatever $\theta_1 \ldots \theta_r$. If this can be done we shall call the region w *similar* to the sample-space W and shall speak of $1 - \alpha$ as its *size*.

The problem of testing composite hypotheses then becomes one of (a) finding the similar regions, and (b) selecting from among those regions the one which minimises the second kind of error for a simple admissible alternative H_t. If this is the same for all H_t we shall have a common best critical region.

26.24. We consider in the first place the composite hypothesis with one degree of freedom. The general problem of finding similar regions in such a case has not been solved, but a solution is possible in one important class of case, namely, that for which

(a) p_0 is indefinitely differentiable with respect to θ_1 for almost all values of θ_1,

(b) the function p_0 obeys the relation

$$\phi' = A + B\phi, \qquad (26.30)$$

where

$$\phi = \frac{\partial}{\partial\theta_1} \log p_0, \qquad \phi' = \frac{\partial\phi}{\partial\theta_1}, \qquad (26.31)$$

and A and B depend on θ_1 but not on the x's. In particular the normal distribution is of this type.

Under conditions (a) and (b) it follows that for w to be similar to W it is necessary and sufficient that

$$\int_w \frac{\partial^k p_0}{\partial\theta_1^k} \, dx = 0, \qquad k = 1, 2, \ldots \qquad (26.32)$$

Let w be a region for which (26.32) is true. Then for $k = 1$ and 2 we have

$$\int_w p_0 \, \phi \, dx = 0$$

$$\int_w p_0 \, (\phi^2 + \phi') \, dx = 0.$$

In virtue of (26.30), this last may be written

$$\int_w p_0 \, (\phi^2 + A + B\phi) \, dx = 0,$$

whence

$$\int_w p_0 \, \phi^2 \, dx = -A \int_w p_0 \, dx = -A \, (1 - \alpha). \qquad (26.33)$$

Differentiating (26.33) with respect to θ_1 and using previous results, we find

$$\int_w p_0\, \phi^3\, dx = (2AB - A')\,(1 - \alpha), \qquad (26.34)$$

and generally

$$\int_w p_0\, \phi^k\, dx = (1 - \alpha)\, \psi_k(\theta_1), \qquad (26.35)$$

where $\psi_k(\theta_1)$ is a function of θ_1 only, and is therefore independent of w. Now (26.32) is true for $W = w$, and we find

$$\int_W p_0\, \phi^k\, dx = \psi_k(\theta_1), \qquad (26.36)$$

so that

$$\frac{1}{1-\alpha}\int_w p_0\, \phi^k\, dx = \int_W p_0\, \phi^k\, dx. \qquad (26.37)$$

Now consider the random variable ϕ. Since p_0 integrated through w is equal to $1 - \alpha$, we may regard $\frac{p_0}{1-\alpha}$ as a frequency function defined in w. It follows from (26.37) that the moments of ϕ in this domain are the same as those of ϕ in W. Consequently, if the moments determine the distribution uniquely, the distributions of ϕ are identical.

Hence we may use the hypersurfaces ϕ = constant to set up similar regions. The space W may be imagined as composed of shells of infinite thinness bounded by these hypersurfaces. If we determine an "area" on one of these shells equal to $1 - \alpha$ times its area in W, the totality of such areas will constitute a region w of size $1 - \alpha$; and since this will be so irrespective of θ_1 the region w is similar to W.

26.25. When similar regions are determined by the above method we have to find the best critical region from among them. Let H_t be a simple admissible alternative. We require to find from the regions w a region w_0 such that

$$\int_{w_0} p_t\, dx = \text{maximum}. \qquad (26.38)$$

We now show that this is equivalent to maximising

$$\int_{w(\phi)} p_t\, dw(\phi), \qquad (26.39)$$

subject to

$$\int_{w(\phi)} p_0\, dw(\phi) = (1 - \alpha)\int_{W(\phi)} p_0\, dW(\phi). \qquad (26.40)$$

Here $w(\phi)$ means the element of w for constant ϕ—the "shell" of the previous section. The object of this is to reduce our present case to that of simple hypotheses. We take ϕ as a new variable and consider together the remaining variables (which amounts to determining similarity of w and W in each separate shell between ϕ and $\phi + d\phi$, as in the previous section), and are thus left with regions dependent on ϕ. Equation (26.39) then requires that the probability of the second kind of error in each shell must be a minimum, subject to the control of the first kind asserted by (26.40).

Suppose that (26.39) were not maximised. There would then exist a set of values of ϕ for each of which we could determine a region $v(\phi)$ such that

$$\int_{v(\phi)} p_0\, dv(\phi) = (1-\alpha)\int_{W(\phi)} p_0\, dW(\phi) \quad . \quad . \quad . \quad (26.41)$$

and

$$\int_{v(\phi)} p_t\, dv(\phi) > \int_{w_0(\phi)} p_t\, dw_0(\phi). \quad . \quad . \quad . \quad . \quad (26.42)$$

Let E be this set of values of ϕ and CE the remaining set. We prove our result by obtaining a contradiction, namely by defining a region v which is similar to W, and such that

$$\int_v p_t\, dx > \int_{w_0} p_t\, dx, \quad . \quad . \quad . \quad . \quad . \quad (26.43)$$

which contradicts (26.38).

Take as v the shells of hypersurfaces (1) in CE which are identical with $w_0(\phi)$ and (2) in E which satisfy (26.42). Now

$$\int_v p_t\, dx = \int_{E+CE} d\phi \int_{v(\phi)} p_t\, dv(\phi)$$

and

$$\int_{w_0} p_t\, dx = \int_{E+CE} d\phi \int_{w_0(\phi)} p_t\, dw_0(\phi).$$

Hence

$$\int_v p_t\, dx - \int_{w_0} p_t\, dx = \int_{E+CE} d\phi \left\{\int_{v(\phi)} p_t\, dv(\phi) - \int_{w_0(\phi)} p_t\, dw_0(\phi)\right\}$$

$$= \int_E d\phi \left\{\int_{v(\phi)} p_t\, dv(\phi) - \int_{w_0(\phi)} p_t\, dw_0(\phi)\right\} > 0, \quad . \quad . \quad (26.44)$$

which is the contradiction required.

26.26. Thus our problem is reduced to that of finding, in the shells ϕ = constant, portions $w_0(\phi)$ which maximise the integral of p_t. We have, so to speak, brought the problem down one dimension by locating it in shells instead of dealing with it throughout the spaces w and W. It now becomes that of a simple hypothesis in $(n-1)$ dimensions, and the best critical region is the one for which

$$p_t \geqslant \frac{1}{k} p_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.45)$$

where k is a function of ϕ. The sum of these regions for the various values of ϕ gives us the complete solution to the problem, and if this sum has boundaries which are independent of H_t we have a common best critical region and a U.M.P. test.

Example 26.6 : "Student's" Hypothesis

A single sample is taken from a normal population

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} dx,$$

with unspecified σ. We have then one degree of freedom, $\theta_1 = \sigma$, and the hypothesis H_0 is that $\mu = \mu_0$, say.

We find

$$\phi = \frac{\partial}{\partial \sigma} \log p_0 = -\frac{n}{\sigma} + \frac{\Sigma (x - \mu_0)^2}{\sigma^3}$$

$$\begin{aligned} \frac{\partial \phi}{\partial \sigma} &= \frac{n}{\sigma^2} - 3\frac{\Sigma (x - \mu_0)^2}{\sigma^4} \\ &= -\frac{2n}{\sigma^2} - \frac{3\phi}{\sigma} \\ &= \frac{n}{\sigma^2} - \frac{3n}{\sigma^4}\{(\bar{x} - \mu_0)^2 + s^2\}. \end{aligned}$$

Condition (26.30) is satisfied, and ϕ is constant over the hypersurfaces

$$\Sigma (x - \mu_0)^2 = n\{(\bar{x} - \mu_0)^2 + s^2\} = \text{constant}.$$

The hypersurfaces are hyperspheres in W. To construct a similar region we have merely to pick out a region of size $1 - \alpha$ on each shell and to amalgamate them. In our present case this is particularly easy because p_0 is constant over the shells and we need only pick out *areas* on each shell bearing to the area of the hypersphere the ratio $1 - \alpha$.

These areas need not be of the same shape or similarly situated. By selecting them in different ways an infinite variety of regions may be constructed. We have to find the best for an alternative simple hypothesis $\sigma = \sigma_1$, $\mu = \mu_1$.

The condition (26.45) becomes

$$\frac{1}{\sigma_1^n} \exp\left[-\frac{n}{2\sigma_1^2}\{(\bar{x} - \mu_1)^2 + s^2\}\right] \geqslant \frac{1}{k\sigma^n} \exp\left[-\frac{n}{2\sigma^2}\{(\bar{x} - \mu_0)^2 + s^2\}\right].$$

As we are dealing with regions which are similar with regard to σ, we may put $\sigma = \sigma_1$ and find

$$\bar{x}(\mu_1 - \mu_0) \geqslant \tfrac{1}{2}(\mu_1^2 - \mu_0^2) - \frac{1}{n}\sigma_1^2 \log k = (\mu_1 - \mu_0) k_1, \text{ say,}$$

where $k_1 = k_1(\phi)$. Thus we find, for the boundary of $w_0(\phi)$,

$$\begin{aligned} &\text{if } \mu_1 > \mu_0, \quad \bar{x} \geqslant k_1(\phi) \\ &\text{if } \mu_1 < \mu_0, \quad \bar{x} \leqslant k_1(\phi), \end{aligned}$$

where k_1 has to be chosen so as to satisfy

$$\int_{w(\phi)} p_0 \, dw(\phi) = (1 - \alpha) \int_{W(\phi)} p_0 \, dW(\phi).$$

Thus on any particular shell the "cap" cut off by the hyperplane $\bar{x} = $ constant must have area $1 - \alpha$ and hence must subtend the same solid angle at the origin. Consequently the boundaries lie on a right hypercircular cone through the point whose co-ordinates are all equal to μ_0 and whose axis is perpendicular to $\bar{x} = 0$, namely the line $x_1 = x_2 = \ldots = x_n$. For each α there will be a different cone. If $\mu_1 > \mu_0$ the cones will be in the positive quadrant and in the contrary case in the negative quadrant.

Furthermore, these regions are independent of μ_1. Thus for the class of hypothesis $\mu_1 > \mu_0$ or $\mu_1 < \mu_0$ (but not both together) the common best critical regions and U.M.P. tests exist.

Finally we have to evaluate α in terms of the sample values determining the critical

cones. We have already seen in Example 10.6 (vol. I, p. 239) that if $z = \frac{x - \mu_0}{s}$ the frequency inside the cone is

$$\frac{1}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \int_0^z \frac{dz}{(1+z^2)^{\frac{n}{2}}} = \alpha.$$

Thus " Student's " test, which we have previously considered on more or less intuitive grounds, is now seen to be the best in the sense of the theory herein developed, for the admissible class $\mu_1 > \mu_0$ or for that $\mu_1 < \mu_0$.

Example 26.7

Consider a sample from the normal population with unspecified mean, the hypothesis being that $\sigma = \sigma_0$. We now find

$$\phi = \frac{\partial}{\partial \mu} \log p_0 = \frac{n(\bar{x} - \mu)}{\sigma_0^2}$$

$$\frac{\partial \phi}{\partial \mu} = -\frac{n}{\sigma_0^2},$$

so that (26.30) is satisfied.

The hypersurfaces ϕ = constant are the hyperplanes $\bar{x}$ = constant, and any regions of size $1 - \alpha$ on these hyperplanes will provide similar regions w. The condition $p_t \geqslant \frac{1}{k} p_0$ will be found to reduce to

$$s^2 (\sigma_0^2 - \sigma_t^2) \leqslant -(\bar{x} - \mu_t)^2 (\sigma_0^2 - \sigma_t^2) + 2\sigma_0^2 \sigma_t^2 \left\{ \log \frac{\sigma_0}{\sigma_t} + \frac{1}{n} \log k \right\} = (\sigma_0^2 - \sigma_t^2) k_1, \text{ say.}$$

If $\sigma_t > \sigma_0$ we have $s^2 \geqslant k_1(\phi)$
and if $\sigma_t < \sigma_0$ we have $s^2 \leqslant k_1(\phi)$.

Since s^2 is independent of $\bar{x}$, k_1 will be a function of α and n only. The best critical regions are those given by $s^2 \geqslant s_0^2$ and $s^2 \leqslant s_0^2$ as the case may be, and the appropriate values of s_0 corresponding to α may be found from the known distribution of s^2. The critical regions are hypercylinders, and again there are two sets of best common critical regions, according as $\sigma_t > \sigma_0$ or $\sigma_t < \sigma_0$.

Composite Hypotheses: Several Degrees of Freedom

26.27. As a preliminary to extending the theory for one degree of freedom to the case of several degrees, we note that if a region w is similar to W with regard to $\theta_1 \ldots \theta_r$ jointly, then it is so for each of them separately; and conversely. The direct result is obvious and the converse follows in this way: (we need prove it only for $r = 2$ because the rest follows step by step). If then

$$\int_w p \, dx = 1 - \alpha$$

is true for $\theta_2, \theta_3 \ldots \theta_r$ independently of θ_1, and for $\theta_1, \theta_3 \ldots \theta_r$ independently of θ_2, then it is true for any values of θ_1 and θ_2 and any other fixed values of $\theta_3 \ldots \theta_r$; and hence it is true independently of θ_1 and θ_2 together.

26.28. An additional preliminary requirement is the concept of independence of a family of surfaces of a parameter. Suppose

$$f_j(x_1 \ldots x_n, \theta) = C_j \qquad j = 1, 2 \ldots k < n \quad . \quad . \quad . \quad (26.46)$$

represents a family of surfaces, where θ and the C's are variable parameters. Let $S(\theta, C_1 \ldots C_k)$ be the intersection of these surfaces, or, if $k = 1$, the surfaces themselves. Consider the family obtained by fixing θ and allowing the C's to vary. Then if any surface of this family for θ_1 can also be obtained from a second family for θ_2 we shall say that the family is independent of θ. We get the same aggregate of intersections however θ is chosen. For example, if

$$f_1 = (x_1 - \theta)^2 + (x_2 - \theta)^2 + (x_3 - \theta)^2 = C_1$$

and

$$f_2 = x_1 + x_2 + x_3 = C_2,$$

the family S consists of circles in planes at right angles to the line $x_1 = x_2 = x_3$ and having their centres on that line. This is true however θ is chosen, and S is therefore independent of θ.

26.29. Under certain restrictive conditions similar to those of **26.24** it is now possible to find solutions to the problem of determining best critical regions. We assume

(1) that $\dfrac{\partial^k p_0}{\partial \theta_j^k}$ exists almost everywhere for all k and $j = 1 \ldots r$;

(2) that if $\quad \phi_j = \dfrac{\partial}{\partial \theta_j} \log p_0 \quad$ and $\quad \phi_j' = \dfrac{\partial \phi_j}{\partial \theta_j}$.

$$\text{then} \quad \phi_j' = A_j + B_j \phi_j; \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.47)$$

(3) that the family of surfaces given by the intersections of $\phi_j = C_j$ is independent of θ_j for $j = 1 \ldots r$.

Subject to these conditions (which are sufficient but not necessary) similar regions exist. Consider any two surfaces ϕ_1 and ϕ_2. Since w is similar with respect to θ_1 alone, we may find surfaces ϕ_1 = constant and

$$\int_{w(\phi_1)} p\, dw(\phi_1) = \int_{W(\phi_1)} p\, dW(\phi_1). \quad . \quad . \quad . \quad (26.48)$$

In accordance with assumption (3), the family of surfaces $\phi_1 = C_1$ is independent of θ_2. Thus if θ_2 varies, $W(\phi_1)$ and $w(\phi_1)$ will not vary, though perhaps they may correspond to other values of C_1. Furthermore, (26.48) is true regardless of θ_2. Hence within the shell ϕ_1 = constant we can repeat the analysis used for one degree of freedom. We find that the necessary and sufficient condition for w to be similar to W with regard to both θ_1 and θ_2 is

$$\int_{w(\phi_1, \phi_2)} p_0\, dw(\phi_1, \phi_2) = (1 - \alpha) \int_{W(\phi_1, \phi_2)} p_0\, dW(\phi_1, \phi_2), \quad . \quad . \quad (26.49)$$

where W is the intersection of $\phi_1 = C_1$, $\phi_2 = C_2$ for any values of C_1 and C_2; and similarly for w.

As before, the most general region w is obtained by amalgamating the portions of size $(1 - \alpha)$ on the intersections of ϕ_1 and ϕ_2. The generalisation to r degrees of freedom is

immediate. It also follows in the usual way that the best critical region is the one for which

$$\int_{w_0} p_t\,dx \geqslant \int_v p_t\,dx,$$

v being any other region of size $1 - \alpha$; and w_0 is defined by

$$p_t \geqslant h(\theta_1 \ldots \theta_r)\,p_0. \quad \ldots \quad (26.50)$$

The following examples will illustrate the theory.

Example 26.8. Ratio of Two Variances

Suppose we have two samples of n_1, n_2 members from independent normal populations whose means and variances are unknown. The joint distribution may be expressed as

$$f \propto \frac{1}{\sigma_1^{n_1}\,\sigma_2^{n_2}} \exp\left[-\frac{n_1}{2\sigma_1^2}\{(\bar{x}_1 - \mu_1)^2 + s_1^2\} - \frac{n_2}{2\sigma_2^2}\{(\bar{x}_2 - \mu_2)^2 + s_2^2\}\right].$$

Consider the composite hypothesis $\sigma_1 = \sigma_2 = \sigma$, say. This has three degrees of freedom, for μ_1, μ_2 and σ are unspecified. As the alternative H_t we will take

$$\theta_1 = \mu_1, \qquad \theta_2 = \mu_2 - \mu_1 = b_1, \qquad \theta_3 = \sigma_1, \qquad \theta_4 = \frac{\sigma_2}{\sigma_1},$$

and for H_0 itself

$$\theta_1 = \mu, \qquad \theta_2 = b, \qquad \theta_3 = \sigma, \qquad \theta_4 = 1.$$

We have first to consider whether the conditions of **26.29** are satisfied.

(1) Evidently p_0 is differentiable for all parameters any number of times.

(2) We find—

$$\phi_1 = \frac{\partial}{\partial\mu}\log p_0 = \frac{1}{\sigma^2}\{n_1(\bar{x}_1 - \mu) + n_2(\bar{x}_2 - \mu - b)\}$$

$$\phi_2 = \frac{\partial}{\partial b}\log p_0 = \frac{n_2}{\sigma^2}(\bar{x}_2 - \mu - b)$$

$$\phi_3 = \frac{\partial}{\partial\sigma}\log p_0 = -\frac{(n_1 + n_2)}{\sigma} + \frac{1}{\sigma^3}\{n_1(\bar{x}_1 - \mu)^2 + n_2(\bar{x}_2 - \mu - b)^2 + n_1 s_1^2 + n_2 s_2^2\}$$

and (26.47) is seen to be satisfied.

(3) The hypersurfaces $\phi_1 = C_1$ are evidently equivalent to

$$n_1\bar{x}_1 + n_2\bar{x}_2 = C_1',$$

where C_1' is an arbitrary parameter. The hypersurfaces $\phi_2 = C_2$ give similarly

$$\bar{x}_2 = C_2'.$$

Both these are independent of θ_2 and their intersections, namely $\bar{x}_1 =$ constant, $\bar{x}_2 =$ constant, are independent of θ_3. Thus the third condition is fulfilled and we may apply the foregoing theory.

The equations $\phi_1 =$ constant, $\phi_2 =$ constant, $\phi_3 =$ constant are equivalent to

$$\begin{aligned} \bar{x}_1 &= \text{constant} \\ \bar{x}_2 &= \text{constant} \\ n_1 s_1^2 + n_2 s_2^2 &= \text{constant} = (n_1 + n_2)\,s_a^2, \text{ say.} \end{aligned}$$

The element w_0 is part of $W(\phi_1, \phi_2, \phi_3)$ within which

$$p_t \geqslant p_0/h\,(\bar{x}_1, \bar{x}_2, s_a)$$

and this condition, by reference to the frequency function, becomes

$$\frac{1}{\sigma^{n_1+n_2}} \exp\left[-\frac{1}{2\sigma^2}\{n_1(\bar{x}_1-\mu)^2+n_1 s_1^2\} - \frac{1}{2\sigma^2}\{n_2(\bar{x}_2-\mu-b)^2+n_2 s_2^2\}\right]$$

$$\leqslant \frac{h}{\sigma_1^{n_1+n_2}\,\theta_4^{n_2}} \exp\left[-\frac{1}{2\sigma_1^2}\{n_1(\bar{x}_1-\mu_1)^2+n_1 s_1^2+n_2\theta_4^{-2}(\bar{x}_2-\mu_1+\mu_2)^2+n_2\theta_4^{-2}s_2^2\}\right].$$

Since the region w is independent of μ, b and σ, we may put them respectively equal to μ_1, b_1 and σ_1 and hence find for the condition

$$n_2(1-\theta_4^2)\{(\bar{x}_2-\mu_1-b_1)+s_2^2\} \leqslant 2\sigma_1^2\,\theta_4^2(\log h - n_2\log\theta_4).$$

Since this inequality holds good on $\bar{x}_2$ = constant it contains only one variable s_2^2 and we accordingly find two cases :—

If $\theta_4 = \dfrac{\sigma_2}{\sigma_1} > 1$ the best region is defined by $s_2^2 \geqslant h_1'(\bar{x}_1, \bar{x}_2, s_a^2)$;

If $\theta_4 = \dfrac{\sigma_2}{\sigma_1} < 1$ the best region is defined by $s_2^2 \leqslant h_2'(\bar{x}_1, \bar{x}_2, s_a^2)$.

We have now to determine h_2' so as to satisfy

$$\int_{w_0(\phi_1,\phi_2,\phi_3)} p_0\,dx = (1-\alpha)\int_{W_0(\phi_1,\phi_2,\phi_3)} p_0\,dx.$$

Now $W(\phi_1, \phi_2, \phi_3)$ is the locus for which $\bar{x}_1$, $\bar{x}_2$ and s_a^2 are constant, and thus the integral on the right is the product of $1-\alpha$ and the frequency function $p_0(\bar{x}_1, \bar{x}_2, s_a^2)$. Similarly that on the left is the integral of this function over the region for which $s_2^2 \gtrless h'$. Thus

$$\int_{w_0} p_0\,dx = \int_{h_1'}^{h_1''} p_0(\bar{x}_1, \bar{x}_2, s_a^2, s_2^2)\,ds_2^2 \text{ in the first case,}$$

with a similar expression but different limits in the second. Now we have for the joint frequency function of $\bar{x}_1$, $\bar{x}_2$, s_1^2 and s_2^2

$$f \propto \frac{1}{\sigma_1^{n_1+n_2}}\, s_1^{n_1-3}\, s_2^{n_2-3} \exp\left[-\frac{1}{2\sigma_1^2}\{n_1(\bar{x}_1-\mu_1)^2+n_2(\bar{x}_2-\mu_2)^2+(n_1+n_2)s_a^2\}\right].$$

Transforming from s_1^2 to s_a^2 as variable, we find for the condition, after a little reduction—

$$\int_{h'}^{h''}\{(n_1+n_2)s_a^2-n_2 s_2^2\}^{\frac{n_1-3}{2}} s_2^{n_2-3}\,ds_2^2 = (1-\alpha)\int_0^{h''}\{(n_1+n_2)s_a^2-n_2 s_2^2\}^{\frac{n_1-3}{2}} s_2^{n_2-3}\,ds_2^2,$$

where $h'' = \dfrac{n_1+n_2}{n_2}s_a^2$. On substituting $n_2 s_2^2 = (n_1+n_2)s_a^2 u$ we find—

$$\int_0^{u_0'}(1-u)^{\frac{n_1-3}{2}} u^{\frac{n_2-3}{2}}\,du = \int_{u_0}^1 (1-u)^{\frac{n_1-3}{2}} u^{\frac{n_2-3}{2}}\,du = (1-\alpha)\,B\left(\frac{n_1-1}{2}, \frac{n_2-1}{2}\right).$$

It follows that u_0, u_0' depend only on α, n_1 and n_2. Thus, whatever the values of $\bar{x}_1$, $\bar{x}_2$ and s_a^2, the best critical region is defined by

$$s_2^2 \geqslant h_1' = \frac{(n_1+n_2)s_a^2}{n_2}u_0 \qquad \text{if } \sigma_2 > \sigma_1$$

$$s_2^2 \leqslant h_2' = \frac{(n_1+n_2)s_a^2}{n_2}u_0' \qquad \text{if } \sigma_2 < \sigma_1.$$

These are equivalent to

$$u = \frac{n_2 s_2^2}{n_1 s_1^2 + n_2 s_2^2} \geqslant u_0 \qquad \text{if } \sigma_2 > \sigma_1$$

$$u \qquad \leqslant u_0 \qquad \text{if } \sigma_1 > \sigma_2.$$

If we put

$$z = \tfrac{1}{2} \log \frac{n_1 (n_2 - 1) s_1^2}{n_2 (n_1 - 1) s_2^2}$$

the B-distribution of u reduces to Fisher's form. The result we have reached is therefore equivalent to showing that the z-test is the best for the ratio of two variances in normal samples. As usual, there is no U.M.P. test for the whole range of the ratio from 0 to ∞, but two U.M.P. tests for the ranges 0 to 1 and 1 to ∞ respectively.

Example 26.9. Difference of Two Means

Consider again the previous example, where now the variances are unspecified but equal and the means μ_1 and $\mu_2 = \mu_1 + b$ may have any values. The hypothesis H_0 is that $b = 0$ and has two degrees of freedom corresponding to μ and σ.

Let the alternative H_t specify the parameters

$$\theta_1 = \mu_t, \qquad \theta_2 = \sigma_t, \qquad \theta_3 = b_t.$$

In addition to the quantities required in the previous Example we now use also $\bar{x}_0$ and s_0^2, the mean and variance of the pooled samples.

We find that the three conditions of **26.29** are satisfied, and

$$\phi_1 = \frac{n_1 + n_2}{\sigma^2} (\bar{x}_0 - \mu_1)$$

$$\phi_2 = -\frac{n_1 + n_2}{\sigma} + \frac{n_1 + n_2}{\sigma^3} \{ (\bar{x}_0 - \mu_1)^2 + s_0^2 \}.$$

Equivalent to this family are the surfaces

$$\bar{x}_0 = C_1$$
$$s_0^2 = C_2.$$

The condition $p_t \geqslant h(\phi_1, \phi_2) p_0$ reduces to

$$b_t (\bar{x}_1 - \bar{x}_2) \leqslant h' (\bar{x}_0, s_0^2),$$

and as usual we find two cases according as $\mu_2 > \mu_1$ or vice-versa. We consider only the first, the second being analogous.

Writing $v = \bar{x}_1 - \bar{x}_2 \geqslant k_2'$ we have to determine h' by

$$\int_{h'''}^{h_1''} p_0 (\bar{x}\ , s_0^2, v)\, dv = (1 - \alpha) \int_{h'''}^{h^{iv}} p_0 (\bar{x}_0, s_0^2, v)\, dv,$$

where h''' and h^{iv} are the lower and upper limits of the variation of v for fixed values of x_0 and s_0^2.

The frequency function of $\bar{x}_0$, s_0^2, v and s_1^2 is easily found to be

$$f \propto s_1^{n_1 - 3} \left\{ (n_1 + n_2) s_0^2 - n_1 s_1^2 - \frac{n_1 n_2}{n_1 + n_2} v^2 \right\}^{\frac{n_2 - 3}{2}} \exp \left[-\frac{n_1 + n_2}{2\sigma^2} \{ (x_0 - \mu_1)^2 + s_0^2 \} \right],$$

whence that of $\bar{x}_0$, s_0^2 and v is found to be

$$f \propto \left(s_0^2 - \frac{n_1 n_2}{(n_1 + n_2)^2} v^2 \right)^{\frac{n_1 + n_2 - 4}{2}} \exp \left[-\frac{n_1 + n_2}{2\sigma^2} \{ (x_0 - \mu_1)^2 + s_0^2 \} \right].$$

Since $\bar{x}_0$ and s_0^2 are constant over the domains under consideration we have to satisfy

$$\int_{h'''}^{h_1''}\left(s_0^2 - \frac{n_1 n_2}{(n_1+n_2)^2} v^2\right)^{\frac{n_1+n_2-4}{2}} dv = 2(1-\alpha)\int_0^{h^{iv}}\left(s_0^2 - \frac{n_1 n_2}{(n_1+n_2)^2} v^2\right)^{\frac{n_1+n_2-4}{2}} dv$$

where

$$h''' = -\frac{(n_1+n_2)\,s_0}{\sqrt{(n_1 n_2)}},\ h^{iv} = \frac{(n_1+n_2)\,s_0}{\sqrt{(n_1 n_2)}}.$$

If we put

$$v = \frac{(n_1+n_2)\,s_0}{\sqrt{(n_1 n_2)}}\,\frac{z}{(1+z^2)^{\frac{1}{2}}},$$

this reduces to

$$\frac{1}{B\left(\frac{1}{2}, \frac{n_1+n_2-2}{2}\right)}\int_{-}^{z_0'}\frac{dz}{(1+z^2)^{\frac{n_1+n_2-1}{2}}} = 1-\alpha$$

and

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 s_1^2 + n_2 s_2^2)}}\sqrt{\frac{n_1 n_2}{n_1+n_2}}.$$

We have thus arrived at the t-test for the difference of two means in normal variation when variances are equal. Once again the test we introduced on more or less intuitive grounds has been shown to be justified in the light of the theory developed in this chapter.

Linear Hypotheses in Normal Variation

26.30. Several of the hypotheses dealt with in foregoing examples are particular cases of a general class known as *linear hypotheses*, which accounts for the fact that we keep arriving at the same sort of conclusions respecting them.

Suppose we have n independent variates typified by x_j distributed in the normal form

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2\sigma^2}(x_j-\mu_j)^2\right\}dx_j$$

with common variance σ^2 but different means. Suppose the means are connected with r and s unknown parameters $\theta_1 \ldots \theta_r \ldots \theta_{r+s}$ by linear equations of the type

$$\mu_k = \sum_j c_{jk}\,\theta_j. \qquad (26.51)$$

Suppose further that the hypothesis H_0 specifies r parameters

$$\theta_1 = B_1, \ldots \theta_r = B_r,$$

and hence is composite with s degrees of freedom. Then H_0 will be called a "linear hypothesis". The reader can verify for himself that "Student's" hypothesis, and the hypothesis as to the difference of two means when variances are equal, are of this type. The homogeneity test in variance-analysis and the test of regression coefficients are also reducible to the same form. If, of course, H_0 specifies r linear relations among the θ's instead of the θ's themselves, it can be reduced to a hypothesis which specifies the θ's directly, except perhaps in degenerate cases which need not detain us.

26.31. The theory developed in the earlier part of the chapter for composite hypotheses may be applied to linear hypotheses as we have defined them, and the argument

follows exactly that of Examples 26.8 and 26.9. It is readily verified that the three conditions of **26.29** are satisfied. We have—

$$\left.\begin{aligned}\phi_j &= \sum_k c_{jk}\frac{x_k-\mu_k}{\sigma^2}\\ \phi_j' &= \text{constant}\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (26.52)$$

$$\left.\begin{aligned}\phi_\sigma &= -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_k (x_k-\mu_k)^2\\ \phi_\sigma' &= -\frac{2n}{\sigma^2}-\frac{3}{\sigma}\phi_\sigma\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (26.53)$$

We can therefore find similar regions $w(\phi_1 \ldots \phi_r, \phi_\sigma)$ and select from them the best critical regions in the usual manner. We will omit the rather cumbrous algebra and quote the following result (Kolodzieczyk, 1935).

Transform to new variates $E_1 \ldots E_{r+s}, y_{r+s+1} \ldots y_n$ by the equation

$$x_k = \mu_k + \sum_{j=1}^{r+s} c_{jk} E_j + \sum_{j=r+s+1}^{n} c_{jk} y_j, \quad . \quad . \quad . \quad . \quad (26.54)$$

where the c's are those given in (26.51) for $j, k \leqslant r+s$ and the other c's are orthogonal, i.e.

$$\left.\begin{aligned}\sum_{i=1}^{k} c_{ki}c_{ji} &= 0, \qquad k \neq j, \qquad j > r+s\\ &= 1, \qquad k = j, \qquad j > r+s\end{aligned}\right\} \quad . \quad . \quad . \quad (26.55)$$

Define

$$nS_a^2 = \sum_{j=r+s+1}^{n} y_j^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.56)$$

and

$$nS_0^2 = \sum_{k=1}^{n}\left(\sum_{j=1}^{r+s} c_{jk}E_j\right)^2. \quad . \quad . \quad . \quad . \quad . \quad (26.57)$$

A further transformation of $E_{r+1} \ldots E_{r+s}$ is now made to variables $\psi_{r+1} \ldots \psi_{r+s}$ so that (26.57) becomes

$$nS_0^2 = \sum_{j,k=1}^{r} R_{jk}E_jE_k + \sum_{k=r+1}^{r+s}\psi_k^2 \quad . \quad . \quad . \quad . \quad (26.58)$$

$$= nS_b^2 + \sum_{k=r+1}^{r+s}\psi_k^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.59)$$

The coefficients R can, of course, be obtained from the c's by ordinary determinantal algebra.

Writing now $\varepsilon_j = \theta_j - \theta_j^0$, i.e. the difference between θ_j on the alternative hypothesis and its value if H_0 is true, we find that the best critical region is given by

$$v = \frac{1}{\sqrt{(nS_a^2+nS_b^2)}}\frac{\sum_{j,k=1}^{r} R_{jk}\varepsilon_j E_k}{\sqrt{\left(\sum_{j,k=1}^{r} R_{jk}\varepsilon_j\varepsilon_k\right)}} \geqslant v_0, \quad . \quad . \quad . \quad (26.60)$$

where v is distributed in the form

$$dF \propto (1 - v^2)^{\frac{n-s-3}{2}} dv \qquad -1 \leqslant v \leqslant 1 \quad . \quad . \quad . \quad (26.61)$$

and v_0 is given by

$$1 - \alpha = \int_{v_0}^{1} dF. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.62)$$

26.32. There is one interesting conclusion to be drawn from (26.60). If a U.M.P. test exists, v should be independent of θ_j and hence of ε_j. This appears to be possible only if the denominator in the second part of (26.60) is rational. But this denominator is seen from (26.59) to have the coefficients of a positive definite form and hence is only rational if $r = 1$. We conclude that if $r \geqslant 2$ no U.M.P. test is possible for linear hypotheses in normal variation.

We have already seen that under general conditions no U.M.P. test exists for $r = 1$. A similar conclusion follows from (26.60) if $r = 1$, for it then becomes

$$\frac{R_{11}\, \varepsilon_1\, E_1}{\sqrt{(R_{11})}\, |\, \varepsilon_1\, |} \geqslant v_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.63)$$

which, as usual, leads to two cases according as $\varepsilon_1 \gtrless 0$.

26.33. We will pause at this point to review our results. We began by defining two kinds of error and showing that a test could be defined as " best " for a single alternative hypothesis if it controlled the first kind and reduced the second to a minimum. When there is a class of admissible alternatives we may sometimes arrive at a U.M.P. test which will minimise errors of the second kind for any member of the class, and such a test may be regarded as the best attainable. Though the U.M.P. test does not exist in the great majority of cases, we may find tests which are U.M.P. for either $\theta_1 > \theta_0$ or $\theta_1 < \theta_0$. Such tests have been reached for " Student's " hypothesis and several others in common use, and are found to give the same tests as those introduced on rather intuitive grounds in Chapter 21.

26.34. The absence of a U.M.P. test implies that in the majority of cases we have to look for other criteria to provide " best " tests. In the remainder of this chapter and in the next we shall consider several lines of approach which have been developed :—

(*a*) Relying on **26.18** we may evolve tests based on the likelihood ratio. These will give U.M.P. tests if such exist, and in the contrary case will do their best, so to speak, by finding the greatest common denominator among the best critical regions.

(*b*) We may consider the properties of tests when the sample number n tends to infinity, and so obtain tests which are U.M.P. in the limit. Such tests, like maximum likelihood estimators, may be employed on the grounds that they are " best " for large n and presumably good for small n.

(*c*) We may derive a new criterion from the concept of bias in statistical tests, which will be explained in the next chapter.

(*d*) Recognizing that there is no test which is U.M.P. everywhere, we may seek for one which is U.M.P. in the neighbourhood of the true value. The idea behind this approach is that it will be more important to detect errors in the neighbourhood of the true value,

and that large errors may be left to look after themselves, either because they are infrequent or because almost any "reasonable" test will reveal them.*

(*e*) When a number of independent parameters are involved, we may abandon the attempt to test for each separately and confine our attention to the class of hypotheses for which they are functionally related, e.g. by $\psi = f(\theta_1 \ldots \theta_r)$. This reduces our problem to the case of a single parameter ψ, and we may be able to show that a particular ψ_0 is the best in the sense that it is U.M.P. with respect to all other ψ's, that is, to all other tests depending on the single function of the unknown parameters.

We proceed to consider these approaches.

Tests Based on Likelihood

26.35. Suppose that for a given member of a composite hypothesis H_0 the joint sampling distribution of the variables $x_1 \ldots x_n$ has a frequency function p_0 (which is, of course, the likelihood). Considering the x's as fixed, we may examine the variation of p_0 according to variation in the unspecified parameters $\theta_1 \ldots \theta_r$ which form a set, say ω. Let p_0 (ω max.) be the maximum value of p_0 for such variation. Similarly, if Ω is the class of admissible alternatives H_1, let p_1 (Ω max.) be the maximum of the likelihood for variations of all the parameters $\theta_1 \ldots \theta_{r+s}$. Write

$$\lambda = \frac{p_0\,(\omega \text{ max.})}{p_1\,(\Omega \text{ max.})}. \qquad (26.64)$$

Then a possible criterion for accepting H_0 is to take as critical regions those points for which

$$\lambda \leqslant \text{constant} = C, \text{ say}, \qquad (26.65)$$

where C is determined by relation to a probability level α from the sampling distribution of λ, which of course is independent of the unknown parameters. In defining λ we have assumed that the maxima on the right of (26.64) exist, but we can give the equation greater generality by taking p_0 (ω max.) as the upper bound of values of p_0 in the set ω where no maximum exists; and so for Ω.

In this form the criterion states that we are to accept H_0 if the maximum likelihood in the set of permissible H_0's is greater than a specified proportion of that in the set of alternatives H_1. In doing so we control the first kind of error in the ordinary way. So far as concerns the second kind of error we saw in **26.18** that for H_0 simple the criterion provided a sort of highest common factor among available tests; and presumably qualities of this kind will be equally useful when H_0 is composite.

The Problem of k Samples

26.36. We will illustrate the theory of the likelihood tests by discussing a problem of considerable practical importance. Suppose we have a sample from each of k normal populations, x_{ij} being the jth member of the ith sample. Let

n_i be the number in the ith sample;
$N = \Sigma(n_i)$ be the total number of observations;
$\bar{x}_i$ be the mean of the ith sample;
s_i^2 be the variance of the ith sample.

* An alternative line would be to concentrate on errors of the second kind for larger deviations, on the ground that large errors are more important than small ones. I understand from Dr. B. L. Welch that he considered this approach shortly before the war; the results did not differ very materially from those given by requiring optimum properties near the true value in the case he examined, and the results were not published.

We will consider three different hypotheses H_0 :—

(1) H, that all populations are the same and hence have the same unspecified mean and unspecified variance.
(2) H_1, that they have the same variance but different unspecified means $\mu_1 \ldots \mu_k$.
(3) H_2, when it is known that they have the same variance, that they have the same means.

We have for the joint likelihood—

$$p = \frac{1}{(2\pi)^{\frac{N}{2}}} \frac{1}{\prod_{i=1}^{k} \sigma_i^{n_i}} \exp\left\{-\sum_{i=1}^{k} n_i \frac{(\bar{x}_i - \mu_i)^2 + s_i^2}{2\sigma_i^2}\right\}.$$

Consider first of all H. We find, for p (Ω max.),

$$\bar{x}_i = \mu_i \qquad (26.66)$$

$$s_i = \sigma_i, \qquad (26.67)$$

and for p (ω max.), putting all the μ's and σ's equal and equating the first partials of $\log p_0$ to zero,

$$\mu_i = \bar{x}_0 = \frac{1}{N}\sum_{i=1}^{k} n_i \bar{x}_i \qquad (26.68)$$

$$\sigma_i^2 = s_0^2 = \frac{1}{N}\sum_{i=1}^{k} n_i \{(\bar{x}_i - \bar{x}_0)^2 + s_i^2\}. \qquad (26.69)$$

Inserting these values in p we find, after a little reduction,

$$\lambda_H = \prod_{i=1}^{k}\left(\frac{s_i^2}{s_0^2}\right)^{\frac{n_i}{2}} \qquad (26.70)$$

Similarly it may be shown that

$$\lambda_{H_1} = \prod_{i=1}^{k}\left(\frac{s_i^2}{s_a^2}\right)^{\frac{n_i}{2}}, \qquad (26.71)$$

where

$$s_a^2 = \frac{1}{N}\sum_{i=1}^{k} n_i s_i^2, \qquad (26.72)$$

and also that

$$\lambda_{H_2} = \left(\frac{s_a^2}{s_0^2}\right)^{\frac{N}{2}}. \qquad (26.73)$$

It will be noticed that $\lambda_H = \lambda_{H_1} \lambda_{H_2}$.

26.37. The function λ_{H_2} may be related to the correlation ratio η^2. We have

$$s_0^2 = s_a^2 + \frac{1}{N}\sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x}_0)^2, \qquad (26.74)$$

and hence

$$\lambda_{H_2} = \left\{1 - \frac{1}{N s_0^2}\Sigma n_i (\bar{x}_i - \bar{x}_0)^2\right\}^{\frac{N}{2}}$$

$$= (1 - \eta^2)^{\frac{N}{2}}. \qquad (26.75)$$

The distribution of λ_{H_2} is thus obtainable directly from the known form for η^2 in samples from an uncorrelated population.

We also find

$$(\lambda_{H_1})^{\frac{2}{N}} = \frac{1}{s_a^2}\{\Pi\,(s_i^2)^{n_i}\}^{\frac{1}{N}} \quad . \quad . \quad . \quad . \quad . \quad (26.76)$$

$$(\lambda_H)^{\frac{2}{N}} = \frac{1}{s_0^2}\{\Pi\,(s_i^2)^{n_i}\}^{\frac{1}{N}}. \quad . \quad . \quad . \quad . \quad . \quad (26.77)$$

The distribution of $(\lambda_{H_2})^{\frac{2}{N}}$ is that of $1 - \eta^2$, where the distribution of η^2 is

$$dF \propto (\eta^2)^{\frac{k-3}{2}}(1-\eta^2)^{\frac{N-k-2}{2}}\,d\eta^2. \quad . \quad . \quad . \quad . \quad (26.78)$$

It can accordingly be tested in this distribution or the related z-form. This is, in fact, the criterion used in the analysis of variance for homogeneity tests, and it is interesting to remark that the z-test here arises in considering the hypothesis that the various distributions parent to the sample values, being already *known* to have the same variance, have the same mean. The other form of hypothesis, H, is that the samples come from the same population, and the equality of variance is not part of the data but part of the hypothesis. We are not then surprised, or should not be so, to find that the λ_H criterion leads to a different test.

26.38. The moments of the distribution of λ_H may be obtained as follows. The joint distribution of $\bar{x}_i$ and s_i is

$$dF \propto \Pi\,(s_i)^{n_i-3}\exp\left[-\Sigma\left\{\frac{n_i}{2\sigma^2}s_i^2 + \frac{n_i}{2\sigma^2}(\bar{x}_i - \mu_i)^2\right\}\right]\Pi\,d\bar{x}_i\,\Pi\,ds_i^2. \quad . \quad (26.79)$$

The distribution of means is independent of that of variances and can be ignored. Further, if

$$\chi^2 = \frac{1}{\sigma^2}\Sigma n_i\,(\bar{x}_i - \bar{x}_0)^2$$

then χ^2 is also independent of the variances, and we have

$$dF \propto \Pi\,(s_i)^{n_i-3}\exp\left(-\Sigma\frac{n_i s_i^2}{2\sigma^2}\right)\chi^{k-2}\exp\left(-\tfrac{1}{2}\chi^2\right)\Pi\,ds_i^2\,d\chi. \quad . \quad . \quad (26.80)$$

Put now

$$\psi_i = \frac{1}{N}\frac{n_i s_i^2}{s_0^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.81)$$

and note that

$$\begin{aligned}\sigma^2\chi^2 &= Ns_0^2 - \Sigma\,n_i s_i^2\\ &= Ns_0^2\,(1 - \Sigma\,\psi_i).\end{aligned} \quad . \quad . \quad . \quad . \quad . \quad (26.82)$$

Transforming to variables ψ and s_0, we find

$$dF \propto \Pi\,\psi_i^{\frac{n_i-3}{2}}(1-\Sigma\psi_i)^{\frac{k-3}{2}}\,\Pi\,d\psi_i\,s_0^{N-3}\exp\left(-\frac{Ns_0^2}{2\sigma^2}\right)ds_0^2,$$

whence, for the distribution of the ψ's,

$$dF \propto \Pi\,\psi_i^{\frac{n_i-3}{2}}(1-\Sigma\psi_i)^{\frac{k-3}{2}}\,\Pi\,d\psi_i. \quad . \quad . \quad . \quad (26.83)$$

Now

$$\lambda_H = \Pi\left(\frac{N\,\psi_i}{n_i}\right)^{\frac{n_i}{2}} \quad . \quad . \quad . \quad . \quad . \quad (26.84)$$

and hence we may find the moments of λ_H by integrating its powers over the distribution

(26.83). Integrals of this kind, known as Dirichlet's, are expressible in terms of gamma functions and we find, for the pth moment of λ_H about zero,

$$\mu_p'(\lambda_H) = \frac{N^{\frac{pN}{2}}\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left\{\frac{(p+1)N-1}{2}\right\}} \prod_1^k \frac{\Gamma\left\{\frac{(p+1)n_i-1}{2}\right\}}{n_i^{\frac{pn_i}{2}}\Gamma\left(\frac{n_i-1}{2}\right)} \quad . \quad . \quad . \quad (26.85)$$

When all the n's are equal this reduces to

$$\mu_p'(\lambda_H) = k^{\frac{pN}{2}}\left\{\frac{\Gamma\left\{\frac{(p+1)n-1}{2}\right\}}{\Gamma\left(\frac{n-1}{2}\right)}\right\}^k \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left\{\frac{(p+1)N-1}{2}\right\}}. \quad . \quad . \quad (26.86)$$

26.39. For the criterion λ_{H_1} we start from the distribution

$$dF \propto \Pi\, s_i^{n_i-3} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(n_i s_i^2)\right\}\Pi\, ds_i^2$$

and on putting

$$\zeta_i = \frac{n_i s_i^2}{N s_a^2} \qquad i = 1, 2 \ldots k-1 \quad . \quad . \quad . \quad . \quad (26.87)$$

$$n_k s_k^2 = N s_a^2\left(1 - \sum_1^{k-1}\zeta_i\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.88)$$

we find, in much the same way as before,

$$dF(\zeta_1 \ldots \zeta_{k-1}) \propto \prod_1^{k-1} \zeta_i^{\frac{n_i-3}{2}}\left(1 - \sum_1^{k-1}\zeta_i\right)^{\frac{n_k-3}{2}}. \quad . \quad . \quad (26.89)$$

Further,

$$\lambda_{H_1} = \left\{\frac{N}{n_k}\left(1 - \sum_1^{k-1}\zeta_i\right)\right\}^{\frac{n_k}{2}} \prod_1^{k-1}\left(\frac{N}{n_i}\zeta_i\right)^{\frac{n_i}{2}} \quad . \quad . \quad . \quad (26.90)$$

whence we find

$$\mu_p'(\lambda_{H_1}) = \frac{N^{\frac{pN}{2}}\Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left\{\frac{(p+1)N-k}{2}\right\}} \prod_1^k \frac{\Gamma\left\{\frac{(p+1)n_i-1}{2}\right\}}{n_i^{\frac{pn_i}{2}}\Gamma\left(\frac{n_i-1}{2}\right)} \quad . \quad . \quad . \quad (26.91)$$

26.40. For large n_i we find, in virtue of the Stirling approximation to the gamma function,

(1) for λ_H $\qquad \mu_p' \to \dfrac{1}{(p+1)^{k-1}}$

(2) for λ_{H_1} $\qquad \mu_p' \to \dfrac{1}{(p+1)^{\frac{k-1}{2}}}$

(3) for λ_{H_2} $\qquad \mu_p' \to \dfrac{1}{(p+1)^{\frac{k-1}{2}}}.$

These limiting forms are the moments of the distributions—

(1) $$\frac{(-\log x)^{k-2}}{\Gamma(k-1)}$$

(2) and (3) $$\frac{(-\log x)^{\frac{k-3}{2}}}{\Gamma\left(\frac{k-1}{2}\right)}.$$

Hence, by the transformation $x = e^{-\frac{1}{2}\chi^2}$ we see that approximately λ_H is distributed as χ^2 with $\nu = 2k-2$, and λ_{H_1} and λ_{H_2} as χ^2 with $\nu = k-1$.

26.41. For small samples Neyman and Pearson have suggested approximating to the distributions of $\lambda_H^{\frac{2}{N}}$ and $\lambda_{H_1}^{\frac{2}{N}}$ by identifying their lower moments with those of the form

$$dF \propto x^{m_1-1}(1-x)^{m_2-1}.$$

This possibility has been examined in detail by Nayer (1936) for the hypothesis H_1 when all the n's are equal. The distribution of λ_H has also been studied by Wilks and Thompson (1937*a*).

26.42. Modified forms of the above tests have been considered by various authors. We may write

$$\log \lambda_{H_1} = \tfrac{1}{2}\Sigma\, n_i \log \frac{s_i^2}{s_a^2}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (26.92)$$

where, of course,

$$s_a^2 = \frac{1}{N}\Sigma\, n_i\, s_i^2.$$

In short, s_a^2 is a weighted mean of the s_i^2 and $(\lambda_{H_1})^{\frac{2}{N}}$ is a weighted geometric mean. Bartlett (1937*c*) has proposed using the degrees of freedom ν_i $(= n_i - 1)$ instead of n_i in these equations, that is to say, defines a criterion

$$\mu^{\frac{2}{\nu}} = \Pi\left(\frac{s_i^2}{\Sigma\, \nu_i\, s_i^2}\right)^{\frac{\nu_i}{\nu}}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (26.93)$$

This test is, in the sense defined in the next chapter, unbiassed, whereas that based on λ_{H_1} is not. Bartlett also suggested as an approximation that $-\dfrac{2\log\mu}{c}$ could be regarded as distributed as χ^2 with $k-1$ degrees of freedom, c being given by

$$c = 1 + \frac{1}{3(k-1)}\left\{\Sigma\left(\frac{1}{\nu_i}\right) - \frac{1}{\nu}\right\}. \qquad . \quad . \quad . \quad . \quad . \quad (26.94)$$

This has recently been reconsidered by Hartley (1940), who showed that it is not very exact for large k and gave a better approximation which can be reduced to tabular form. Cf. Exercise 27.2.

Likelihood Criteria for the Linear Hypothesis

26.43. We now proceed to consider the application of the likelihood criterion to the class of linear hypothesis as defined in **26.30.** We have, for the likelihood function,

$$p_0 = \left(\frac{1}{\sigma\sqrt{(2\pi)}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\Sigma(x_j-\mu_j)^2\right\}. \qquad (26.95)$$

Writing $S^2 = \Sigma(x_j-\mu_j)^2$ we have, for the stationary values of p_0 with respect to σ and the parameters θ (related to the μ's by (26.51)),

$$\frac{\partial}{\partial\sigma}\log p_0 = -\frac{n}{\sigma}+\frac{S^2}{\sigma^3} = 0 \qquad (26.96)$$

$$\frac{\partial}{\partial\theta_j}\log p_0 = \frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-\mu_k)\,c_{jk} = 0. \qquad (26.97)$$

This last equation is clearly the one we should get if we were seeking to minimise S^2 itself for variations in the θ's. Let nS_a^2 be this minimum value. We shall then have, from (26.96),

$$\sigma^2 = S_a^2. \qquad (26.98)$$

The maximum of p in the class Ω of admissible hypotheses is then

$$p\,(\Omega\text{ max.}) = \left(\frac{1}{S_a\sqrt{(2\pi)}}\right)^n e^{-\frac{n}{2}}. \qquad (26.99)$$

Similarly the maximum of p in the class ω for which $\theta_1 \ldots \theta_r$ are fixed and the other s θ's vary, is found to be

$$p\,(\omega\text{ max.}) = \left(\frac{1}{\sqrt{(S_a^2+S_b^2)}\,\sqrt{(2\pi)}}\right)^n e^{-\frac{n}{2}}, \qquad (26.100)$$

where $n(S_a^2+S_b^2)$ is the minimum of S^2 under the conditions that $\theta_1 \ldots \theta_r$ are fixed. Thus we find for the likelihood ratio λ

$$\lambda^{\frac{2}{n}} = \frac{1}{\left(1+\dfrac{S_b^2}{S_a^2}\right)}, \qquad (26.101)$$

or, if more convenient, we may use the function

$$Z = \frac{S_b}{S_a}$$

to provide a criterion.

Now we make the transformation (26.54) and show that the values S_a and S_b as we have defined them here have, in fact, the values given by (26.56) and (26.59). We have, from (26.54),

$$S^2 = \Sigma(x_j-\mu_j)^2 = \sum_{k=1}^{n}\left\{\sum_{j=1}^{r+s}c_{jk}E_j + \sum_{j=r+s+1}^{n}c_{jk}\,y_j\right\}^2$$

$$= \sum_{k=1}^{n}(\Sigma\, c_{jk}E_j)^2 + \sum_{k=1}^{n}(\Sigma\, c_{jk}\,y_j)^2$$

$$= \sum_{k=1}^{n}(\Sigma\, c_{jk}E_j)^2 + \sum_{j=r+s+1}^{n} y_j^2.$$

Since $n S_a^2$ is the minimum of S^2 for all variations of the θ's and E and y are independent of the θ's, we must have

$$nS_a^2 = \Sigma\, y_j^2.$$

Also, since nS_b^2 is the minimum of S^2 when the values $\theta_1 \ldots \theta_r$ are fixed, it is seen to have the value given in (26.59).

We have also

$$S^2 = nS_a^2 + nS_0^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.102)$$

where

$$nS_0^2 = \sum_{k=1}^{n} \left(\sum_{j=1}^{r+s} c_{jk}\, E_j \right)^2,$$

and the frequency function of E's and y's is given by

$$f(E_1 \ldots E_{r+s}, y_{r+s+1} \ldots y_n) \propto \exp\left\{ -\frac{n}{2\sigma^2}(S_a^2 + S_0^2) \right\}. \quad . \quad (26.103)$$

Now nS_a^2 is the sum of squares of $n - r - s$ normal variates, and hence

$$f(S_a) \propto S_a^{n-r-s-1} \exp\left(-\frac{nS_a^2}{2\sigma^2} \right). \quad . \quad . \quad . \quad (26.104)$$

Hence, since the E's are independent of the y's, and since S_a^2 depends only on the y's,

$$f(S_a, E_1 \ldots E_{r+s}) \propto S_a^{n-r-s-1} \exp\left\{ -\frac{n}{2\sigma^2}(S_a^2 + S_0^2) \right\}. \quad . \quad (26.105)$$

We have seen, in effect, that $n\, S_b^2$ is the minimum value of S_0^2. It depends on $E_1 \ldots E_r$ and hence is independent of S_a^2 and is distributed as

$$f(S_b) \propto S_b^{r-1} \exp\left(-\frac{nS_b^2}{2\sigma^2} \right).$$

Thus we have

$$f(S_a, S_b) \propto S_a^{n-r-s-1} S_b^{r-1} \exp\left\{ -\frac{n}{2\sigma^2}(S_a^2 + S_b^2) \right\}. \quad . \quad . \quad . \quad (26.106)$$

Putting now $Z = S_b/S_a$, we find

$$f(Z) \propto Z^{r+s-1}(1 + Z^2)^{-\frac{n-s}{2}} \quad . \quad . \quad . \quad . \quad . \quad (26.107)$$

which may be reduced to Fisher's form by putting

$$z = \tfrac{1}{2} \log \frac{S_b^2(n - r - s)}{rS_a^2} = \log Z + \tfrac{1}{2} \log \frac{n - r - s}{r}. \quad . \quad . \quad (26.108)$$

We have thus reduced the test of the linear hypothesis to the z-test and it is seen that several of the tests introduced in Chapter 21 can be justified on the likelihood criterion. These include the "Student" test for one mean, the extended form for the difference of two means, and the test for the ratio of variances. Certain other tests in which the z-distribution (which, of course, reduces to the t-distribution for $\nu_1 = 1$) appears—such as that of the correlation ratio, the multiple correlation coefficient and regression coefficients—also depend on the linear hypotheses, and in the light of the theory here presented are seen to be different aspects of the same thing, at least so far as the testing of hypotheses is concerned.

26.44. We will indicate briefly, without going into the complicated mathematics involved, some interesting results obtained by P. C. Tang (1938) and P. L. Hsu (1941*b*) concerning the power of the z-test as applied to linear hypotheses.

The functions S_a^2 and S_b^2, as we have seen, are distributed independently in the χ^2-form, and their ratio accordingly in Fisher's form. From this viewpoint the test of the linear hypothesis is a generalisation of the test of homogeneity in the analysis of variance. Tang considers the distribution of

$$E^2 = \frac{S_b^2}{S_a^2 + S_b^2} = 1 - \lambda^{\frac{2}{n}} \quad . \quad . \quad . \quad . \quad (26.109)$$

and the variation for errors of the second kind, namely, when the values $\theta_1 \ldots \theta_r$ are different from the specified values. He shows that the power of the test depends, not on individual alternative values, but on a single function of the θ's. He also obtains the power function and tabulates it.

Hsu then considers other possible tests which are based on this single function and shows that in this class of test the z-test or the equivalent E^2-test is the uniformly most powerful.

26.45. For large samples, when maximum likelihood estimators of the parameters exist, the distribution of $-2\log\lambda$ is that of χ^2 with s degrees of freedom. For the distribution may then be written (see **17.46**)—

$$dF = A \exp\left\{-\frac{n}{2}\Sigma\, g_{jk}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)\right\} d\hat{\theta}_1 \ldots d\hat{\theta}_{r+s}$$

so that

$$p\,(\Omega \text{ max.}) = A. \quad . \quad . \quad . \quad . \quad . \quad (26.110)$$

If $\theta_1 \ldots \theta_r$ are fixed the likelihood becomes

$$p = A \exp\left\{-\frac{n}{2}\Sigma\, g'_{jk}\, z'_j\, z'_k - \tfrac{1}{2}\chi_0^2\right\},$$

where

$$\chi_0^2 = \sum_{j,\,k=1}^{r} g'_{jk}(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) \quad . \quad . \quad . \quad . \quad (26.111)$$

and z'_j is given by $\hat{\theta}_j - \theta_j - L_j$ where L_j is a linear function of the r specified parameters. Thus—

$$p\,(\omega \text{ max.}) = A_0\, e^{-\frac{1}{2}\chi_0^2}, \quad . \quad . \quad . \quad . \quad . \quad (26.112)$$

where A_0 is the value of A when θ_j takes its true value θ_{j0}. Thus, when H_0 is true,

$$\lambda = e^{-\frac{1}{2}\chi_0^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.113)$$

But the characteristic function of χ_0^2 ($= -2\log\lambda$) is

$$\int p_0\, e^{it\chi_0^2}\, d\hat{\theta}_1 \ldots d\hat{\theta}_{r+s}$$

$$= A\int \exp\left\{-\frac{n}{2}\Sigma\, g'_{jk}\, z'_j\, z'_k + \chi_0^2(it - \tfrac{1}{2})\right\} d\hat{\theta}_1 \ldots d\hat{\theta}_{r+s}$$

$$\propto \frac{1}{(1-2it)^{\frac{s}{2}}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.114)$$

This is the characteristic function of a quantity distributed as χ^2 with s degrees of freedom, and hence the result follows.

26.46. In concluding this chapter we may mention briefly a question which frequently presents itself when statistical hypotheses are being tested in practice. Our tests are based on the observed values obtained in the sampling process, and in order to apply

them we require no prior knowledge of the parameters to which they relate. They can be used in a state of complete ignorance about the parameters. But suppose some information is already available; or suppose that we attach varying degrees of importance to the avoidance of particular types of error. How far are the tests developed in this chapter to be modified?

26.47. Consider, for example, the situation which has already been mentioned in connection with the theory of estimation, of the chemist who is assaying the strength of a particular drug. If the drug has harmful effects in large quantities it may be much more important for him to detect cases in which the true strength exceeds his hypothetical value than when the true strength is deficient. Again, the manufacturer of a "guaranteed" product is usually much more concerned with ensuring that it does not fall below the guaranteed standard than that it exceeds such standard. In such circumstances we may be particularly interested in "one-sided" tests of the type $\xi \leqslant \xi_0$, and as we have seen, there more often occur U.M.P. tests for this class of alternative than in the case when ξ can have any value. We might, therefore, be quite ready to accept such a test, knowing quite well that it may be insensitive in part of the range of the unknown parameter, merely because errors in that range are relatively unimportant.

Similarly we might be willing to accept a test which had a poor discriminatory power in part of the range but compensating advantages elsewhere, simply because we know beforehand that values of the parameter rarely or never fall into that particular part of the range. This is equivalent to prior knowledge of the distribution of the values determining the alternative hypotheses.

26.48. It is difficult to reduce rather vague prior knowledge of a parameter to numerical form, and hence to extend our theory with great precision to cover these cases; but in practice it is desirable to consider, before adopting a test, whether any prior knowledge is available, or whether our interests centre on particular parts of the range. If they do, we may consider the behaviour of power functions of the possible tests at our disposal and examine which is the more powerful test *in the particular part of the range which interests us most.* The mere fact that the theory developed in this and the succeeding chapter makes no assumptions about the prior probabilities of admissible alternatives does not mean that we should be acting sensibly in ignoring any prior information which may be at hand when applying the theory, or that we need feel compelled to apply tests with optimum properties in regions where we know the unknown parameter-values will not fall.

NOTES AND REFERENCES

The theory of this chapter is very largely due to Neyman and E. S. Pearson, whose treatment has been closely followed. In their first contribution to the subject (1928) the likelihood criterion was developed, the theory of first and second kind of errors and power of tests being given in 1933. For the theory of unbiassed tests, see the papers of 1936 and 1938. In the last few years the literature has grown considerably.

Feller (1938) has shown that similar regions only exist in rather exceptional circumstances and that the theory of composite hypotheses is incomplete. Tables of certain power functions and distributions associated with likelihood tests are given by Mahalanobis (1933), Neyman and Tokarska (1936*b*), Wilks and Thompson (1937*a*), P. C. Tang (1938),

David (1939), Nayer (1936), and in *Tables for Statisticians*, Part II (Tables 35–37). See also Mahalanobis (1933).

For tests based on the likelihood ratio, see Neyman and Pearson (1928, 1931*a*, 1931*b*), Pearson and Wilks (1933*b*), Wilks (1935*a*), Nayer (1936), Welch (1936*a*), R. W. Jackson (1936), Sukhatme (1936*b*), Bartlett (1937*c*), Wilks and Thompson (1937*a*), Wilks (1938*a*), Bishop (1939), G. W. Brown (1939), Mood (1939), Hartley (1940), Wald and Brookner (1941*b*).

For the general theory, see also Welch (1935), Kolodzieczyk (1935), Neyman (1935*b*, 1937*b*, 1938*b*), Daly (1940), Pitman (1939*b*), Wald (1939*a*, 1941*a*), Wolfowitz (1942), E. S. Pearson (1941, 1942*a*), Dantzig (1940), P. L. Hsu (1941*b*), Simaika (1941), MacStewart (1941), Scheffé (1942*a*, 1943).

EXERCISES

26.1. Examine the following argument: To accept H when it is false is equivalent to rejecting not-H when not-H is true. Hence, if K = not-H, to commit an error of the second kind for H is to commit an error of the first kind for K; and thus there is no distinction between the first and second kinds of error.

26.2. For the distribution

$$\begin{aligned} dF &= \beta\, e^{-\beta(x-\gamma)}\, dx, && x \geqslant \gamma \\ &= 0 && x < \gamma \end{aligned}$$

show that for a hypothesis H_0 that $\beta = \beta_0$, $\gamma = \gamma_0$ and an alternative H_1 that $\beta = \beta_1$, $\gamma = \gamma_1$, the best critical region is the region W_0 where $p_0 = 0$, together with the region W_+ defined by

$$\bar{x} \leqslant \frac{1}{\beta_1 - \beta_0}\left\{\gamma_1\beta_1 - \gamma_0\beta_0 - \frac{1}{n}\log k + \log\frac{\beta_1}{\beta_0}\right\},$$

provided that the admissible hypothesis is restricted by the conditions $\gamma_1 \leqslant \gamma_0$, $\beta_1 > \beta_0$. Hence show that a U.M.P. test exists in such circumstances.

(Neyman and Pearson, 1936*a*. This shows that a U.M.P. test can exist for more than one unknown parameter.)

26.3. If the distribution function of $x_1 \ldots x_n$ is given by

$$dF = \frac{1}{\sigma^n(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{j=1}^{n} x_j - n\gamma\right)^2 - \tfrac{1}{2}\sum_{j=2}^{n} x_j^2\right\} dx_1 \ldots dx_n,$$

$$\gamma,\ \sigma > 0, \qquad -\infty \leqslant x_1 \ldots x_n \leqslant \infty$$

show that the frequency function may be put in the form

$$f \propto \exp\left(-\frac{n^2(\bar{x}-\gamma)^2}{2\sigma^2}\right)\exp\left(-\tfrac{1}{2}\sum_{2}^{n} x_j^2\right);$$

and hence that $\bar{x}$ is a "shared" estimator sufficient for γ and σ. Show further that the best critical regions for γ_0, σ_0 differ according as $\sigma^2 > \sigma_0^2$, $\sigma^2 < \sigma_0^2$ or $\sigma = \sigma_0$, and that their boundaries depend on γ. Hence no U.M.P. test exists for admissible alternatives $\sigma > 0$.

(Neyman and Pearson, 1936*a*.)

26.4. In the previous exercise put $\sigma = \gamma$ and consider the class of hypothesis $\gamma > 0$. Show that there are different best critical regions according as $\gamma > \gamma_0$, $\gamma < \gamma_0$ and that their boundaries depend on γ. Hence there is no U.M.P. test, but $\bar{x}$ is sufficient for γ.

(Neyman and Pearson, 1936*a*.)

26.5. In samples from a normal population, show that the probability of accepting the hypothesis that the mean $\mu \leqslant \mu_0$ when, in fact, it is false and $\mu = \mu_1 > \mu_0$—that is, the probability of an error of the second kind—is

$$\left(\frac{n}{t}\right)^n \frac{1}{2^{\frac{1}{2}(n-1)}\Gamma(\frac{1}{2}n)} \int_0^\infty v^{n-1} \exp\left(-\frac{nv^2}{2t^2}\right) \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{v-\rho} e^{-\frac{1}{2}u^2}\, du\, dv$$

where
$$\rho = \frac{\mu_1 - \mu_0}{\sigma}$$

and t is the value of $\dfrac{\bar{x} - \mu}{s}$ corresponding to the significance level $1 - \alpha$ for the control of errors of the first kind.

(Neyman and Tokarska, 1936*b*.)

26.6. In six samples of six members each the following values were obtained—

Sample.	Mean.	s_i^2.
1	8433	24,722
2	8200	94,133
3	7933	149,733
4	8120	45,037
5	7971	88,480
6	8263	49,921

with $s_0^2 = 104{,}588$, $S_a^2 = 75{,}338$.

Show that $\lambda_{H_1}^{\frac{2}{N}} = 0{\cdot}8508$ and $\lambda_H^{\frac{2}{N}} = 0{\cdot}6219$. The 5-per-cent. levels are respectively 0·67 and 0·54, so that there is no evidence of heterogeneity.

(Pearson, appendix to papers by Wilsdon, 1934).

26.7. Verify that the likelihood ratio leads to "Student's" test for an unknown mean in normal samples, to the use of Fisher's z in testing the equality of two variances, and to the t-test for the difference of two means in normal populations with the same variance.

26.8. If samples $n_1 \ldots n_k$ are drawn from the populations

$$dF = \frac{1}{\sigma_i} \exp\left(-\frac{x - \beta_i}{\sigma_i}\right) dx, \qquad i = 1 \ldots k$$

use the likelihood ratio to test the hypothesis H_0 that the populations are identical, showing that

$$L_0^N = \lambda_{H_0} = \frac{\prod\limits_{i=1}^{k} (\bar{x}_i - x_{i1})^{n_i}}{(\bar{x}_0 - x_{.1})^N} = \frac{\Pi\, l_i^{n_i}}{l_0^N}, \text{ say,}$$

where $\bar{x}_i$ is the mean of the ith sample, x_{i1} is the smallest member of that sample, $\bar{x}_0$ is the mean of all samples together and $x_{.1}$ is the smallest value in all samples together.

Show that the distribution of x_{i1} and l_i is

$$f = \frac{1}{\sigma^{n_i}}\left(\frac{n_i^{n_i}}{(n_i - 2)!}\right) l_i^{n_i - 2} \exp\left\{-\frac{n_i(l_i + x_{i1})}{\sigma}\right\}$$

and hence the moments of L_0 are

$$\mu_p' = \frac{N^p\,\Gamma(N-1)}{\Gamma(N+p-1)} \prod_{i=1}^{k} \left\{\frac{\Gamma\left(n_i - 1 + \frac{pn_i}{N}\right)}{n_i^{\frac{pn_i}{N}}\,\Gamma(n_i - 1)}\right\}.$$

If H_1 is the hypothesis that the populations have the same σ but any possible different β's, show that

$$L_1^N = \lambda_{H_1} = \frac{\Pi\, l_i^{n_i}}{\bar{l}^N},$$

where $\bar{l}$ is the weighted mean of the l's, and that

$$\mu_p(L_1) = \frac{N^p\,\Gamma(N-k)}{\Gamma(N-k+p)}\,\Pi \left\{\frac{\Gamma\left(n_i - 1 + \frac{pn_i}{N}\right)}{n_i^{\frac{pn_i}{N}}\,\Gamma(n_i - 1)}\right\}.$$

If H_2 is the hypothesis that the populations, being known to have identical σ's, have the same β, show that the distribution of

$$L_2 = \lambda_{H_2}^{\frac{1}{N}} = \frac{\bar{l}}{l_0}$$

is

$$dF = \frac{\Gamma(N-1)}{\Gamma(N-k)\,\Gamma(k-1)}\, L_2^{N-k-1}(1 - L_2)^{k-2}\, dL_2.$$

(Sukhatme, 1936*b*).

26.9. In the notation of **26.36** show that, if H is true, the criteria λ_{H_1} and λ_{H_2} are distributed independently.

(Neyman and Pearson, 1931*b*).

GENERAL THEORY OF SIGNIFICANCE-TESTS—(2)

Bias in Statistical Tests

27.1. In considering the problem of estimation by confidence intervals in Chapter 19 we had occasion to remark on the rather arbitrary nature of determining the interval so that both inequalities $\theta_1 \leqslant \theta$ and $\theta \leqslant \theta_2$ had an equal chance $\frac{1}{2}\alpha$ of fulfilment. A point of a similar nature arises in the testing of hypotheses, particularly when an asymmetrical sampling distribution for the criterion is concerned. Consider, for instance, the testing of the hypothesis that in a normal sample of n members the standard deviation σ has an assigned value σ_0 irrespective of the mean μ. As we have seen in Example 26.3, there is no U.M.P. test for all $\sigma > 0$, though there is one for $\sigma > \sigma_0$ and another for $\sigma < \sigma_0$. In choosing a test to cover the whole range $\sigma > 0$ we have, therefore, a certain freedom of choice, since there exists no "best" test as we have previously defined the term. A common test in practical use is to take the sample variance s^2 and accept the hypothesis $\sigma = \sigma_0$ if and only if

$$s_1^2 \leqslant s^2 \leqslant s_2^2, \quad \ldots \quad (27.1)$$

where s_1^2 and s_2^2 are determined from the distribution of s^2, namely

$$dF \propto s^{n-3} \exp\left(-\frac{ns^2}{2\sigma_0^2}\right) d(s^2), \quad \ldots \quad (27.2)$$

such that

$$\int_0^{s_1^2} dF = \int_{s_2^2}^{\infty} dF = \tfrac{1}{2}(1-\alpha). \quad \ldots \quad (27.3)$$

In short, s_1^2 and s_2^2 are chosen so as to cut off equal "tail" areas of the distribution. This procedure will, of course, control errors of the first kind; but so equally well would the selection of s_1^2 and s_2^2 so that

$$\int_0^{s_1^2} dF = \tfrac{1}{2} - \alpha_1 \quad \ldots \quad (27.4)$$

and

$$\int_{s_2^2}^{\infty} dF = \tfrac{1}{2} - \alpha_2, \quad \ldots \quad (27.5)$$

provided that $\alpha_1 + \alpha_2 = \alpha$. Thus we have an infinite number of regions which will control errors of the first kind. It is natural to seek for some criterion which will distinguish one as better than the others, recognizing that no U.M.P. test exists.

27.2. Such a criterion arises naturally from the following consideration. In the example given, with $\alpha_1 = \alpha_2 = \frac{1}{2}\alpha$, let us calculate the power of the test for different values of σ. This can readily be done from the distributions of type (27.2) by means of the incomplete Γ-function or the equivalent χ^2 integral. For any given σ we have to find

$$\beta(s_1^2, s_2^2 \mid \sigma) = \int_0^{s_1^2} + \int_{s_2^2}^{\infty} dF, \quad \ldots \quad (27.6)$$

where

$$dF = \frac{\left(\frac{n}{2}\right)^{\frac{1}{2}n-1}}{\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{s^2}{\sigma^2}\right)^{\frac{n-3}{2}} e^{-\frac{ns^2}{2\sigma^2}} d\left(\frac{s^2}{\sigma^2}\right). \qquad (27.7)$$

Fig. 27.1, adapted from Neyman and Pearson (1936), shows the relation between the power function β and σ^2 for $\alpha_1 = \alpha_2 = 0{\cdot}49$, $n = 3$, the rejection level being $0{\cdot}02$.

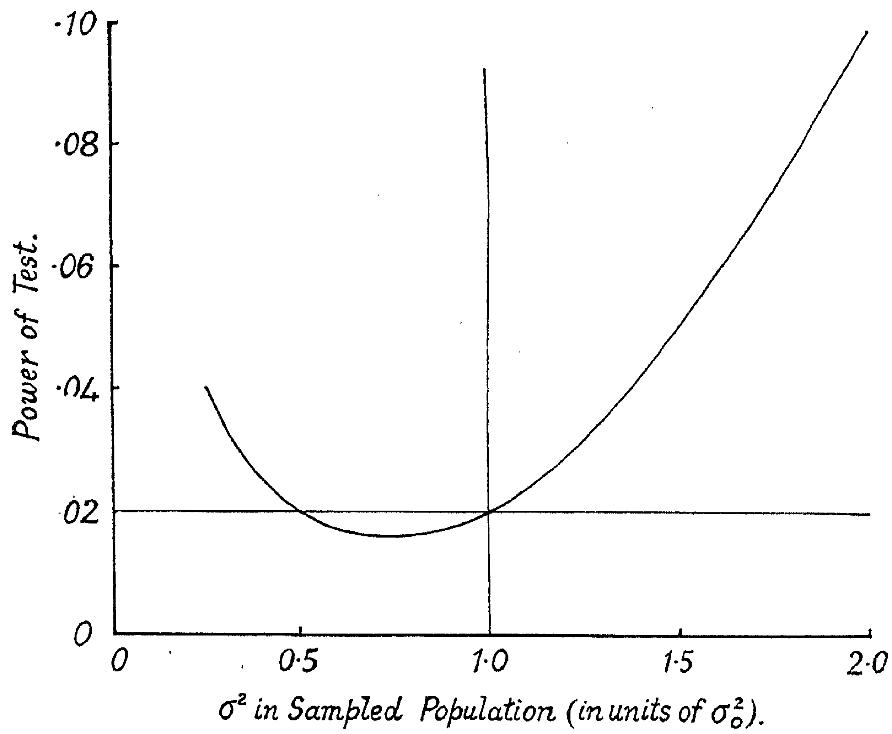


FIG. 27.1.—Power Curve in Samples of 3 for σ^2 from a Normal Population (see text).

We see that for $\sigma > 1 = \sigma_0$ the power increases, and so also for $\sigma < \frac{1}{2} = \frac{1}{2}\sigma_0$. But between $\frac{1}{2}\sigma_0$ and σ_0 the power is less than $0{\cdot}02$, i.e. less than $1 - \alpha$. Hence for such values the chance of an error of the second kind, namely, the acceptance of a false hypothesis, would be greater than the chance of an error of the first kind, namely, the rejection of a true hypothesis.

27.3. Whether this is felt to be anomalous depends on the relative importance of the two kinds of error in particular cases; but, other things being equal, it may be felt more important to avoid the second kind than the first, and not to have a greater probability of accepting the hypothesis when it is false than of rejecting it when it is true. This, at any rate, is the basis of the criterion which we proceed to discuss, namely, that the critical region w should be chosen so that $P(E \varepsilon w)$ is a minimum when the hypothesis tested is true.

Consider then the case when H_0 ascribes to a parameter θ the value θ_0, and the admissible alternatives ascribe other values to θ but do not differ from H_0 in other respects. We shall say that w is an *unbiassed* critical region if, and only if,

$$\int_w p_0\, dx = P(E \varepsilon w \mid \theta_0) = 1 - \alpha, \qquad (27.8)$$

and for any other θ, say θ',

$$\int_w p(\theta')\, dx = P(E \varepsilon w \mid \theta') \geqslant 1 - \alpha. \qquad (27.9)$$

Equation (27.8) expresses the usual control of errors of the first kind and (27.9) the minimising property of w. If a region is not unbiassed it will be said to be biassed.

27.4. In certain cases there will exist among the unbiassed regions a w_0 such that

$$\int_{w_0} p(\theta')\,dx \geqslant \int_{w} p(\theta')\,dx \quad . \quad . \quad . \quad . \quad . \quad (27.10)$$

for all admissible θ'. Such a region may be called the best unbiassed critical region and the test based on it the uniformly most powerful unbiassed test, or briefly the U.M.P.U. test. It minimises the risk of errors of the second kind among the class of unbiassed tests. As we shall see presently, U.M.P.U. tests do in fact exist in certain cases.

The use of the word " unbiassed " in this connection is rather arbitrary and is not to be interpreted as meaning that biassed tests will give systematically wrong results, or that unbiassed tests are based on unbiassed estimators. Fortunately the different uses of the term " bias " usually occur in different contexts and confusion is infrequent.

Unbiassed Regions of Type A

27.5. Following Neyman and Pearson, we now define an unbiassed critical region of Type A as one for which

$$\int_{w} p_0\,dx = 1 - \alpha, \quad . \quad . \quad . \quad . \quad (27.11)$$

$$\left[\frac{\partial}{\partial\theta}\int_{w} p\,dx\right]_{\theta=\theta_0} = 0, \quad . \quad . \quad . \quad . \quad . \quad (27.12)$$

and

$$\left[\frac{\partial^2}{\partial\theta^2}\int_{w} p\,dx\right]_{\theta=\theta_0} \text{ is a maximum.} \quad . \quad . \quad . \quad (27.13)$$

We shall, as usual, assume that the differential coefficients exist and shall also assume that differentiation may be carried out under the integral sign, so that we have for all w,

$$\frac{\partial}{\partial\theta}\int_{w} p\,dx = \int_{w}\frac{\partial p}{\partial\theta}\,dx = \int_{w} p'\,dx, \text{ say,} \quad . \quad . \quad . \quad (27.14)$$

and similarly for the second differential coefficient which we denote by p''.

The first condition (27.11) controls errors of the first kind; the second makes the region w locally unbiassed; the third, (27.13), implies that as θ departs from θ_0 the power function increases more rapidly than for any other unbiassed critical region of the same size. Thus *in the neighbourhood of* θ_0 the test may be said to be better than others of the unbiassed type. It may not be better for larger values of $|\theta - \theta_0|$, but the Type A tests are based on the supposition that it is more important to detect small errors of the second kind than to minimise the risk of large errors, which will probably be detected in any case.

27.6. The regions of Type A may be found by the use of the following theorem: the region w_0 is an unbiassed critical region of Type A if, within w_0,

$$p''(\theta_0) \geqslant k_1 p'(\theta_0) + k_2 p(\theta_0), \quad . \quad . \quad . \quad . \quad (27.15)$$

and outside w_0,

$$p''(\theta_0) \leqslant k_1 p'(\theta_0) + k_2 p(\theta_0), \quad . \quad . \quad . \quad . \quad (27.16)$$

where

$$p'(\theta_0) = \left[\frac{\partial p}{\partial\theta}\right]_{\theta=\theta_0} \text{ etc.,}$$

and k_1, k_2 are chosen so as to satisfy (27.12) and (27.13).

Suppose that $F_0 \dots F_m$ are functions of $x_1 \dots x_n$ and that

$$\int_w F_j \, dx = c_j, \quad \text{a constant.} \qquad (27.17)$$

Let w_0 be a region such that inside it

$$F_0 \geqslant \sum_{j=1}^{m} k_j F_j \, . \qquad (27.18)$$

and outside it

$$F_0 \leqslant \Sigma k_j F_j, \qquad (27.19)$$

where the k's are constants chosen so as to satisfy (27.17). Then for any w for which (27.17) is valid

$$\int_w F_0 \, dx \leqslant \int_{w_0} F_0 \, dx. \qquad (27.20)$$

In fact, let ww_0 be the common part, if any, of w and w_0. As both w and w_0 satisfy (27.17), we have

$$\int_{w-ww_0} F_j \, dx = \int_{w_0-ww_0} F_j \, dx. \qquad (27.21)$$

Now

$$\int_{w_0} F_0 \, dx - \int_w F_0 \, dx = \int_{w_0-ww_0} F_0 \, dx - \int_{w-ww_0} F_0 \, dx$$

$$\geqslant \int_{w_0-ww_0} \Sigma (k_j F_j) \, dx - \int_{w-ww_0} \Sigma (k_j F_j) \, dx$$

$$\geqslant 0,$$

in virtue of (27.21).

In our present case take F_0 as $p''(\theta_0)$ and F_1, F_2 as $p'(\theta_0)$, $p(\theta_0)$ respectively. Then (27.20) is true, and hence (27.13) is satisfied if (27.18) and (27.19) are true; and these will be found to reduce to conditions (27.15) and (27.16). The theorem follows.

27.7. If (27.14) holds, and if there exists a sufficient estimator t for θ, then the Type A region is bounded by surfaces of constant t. For then we have

$$p(\theta) = p_1(t, \theta)\, p_2(x) \qquad (27.22)$$

and hence, from (27.15), on substitution,

$$p_1''(t, \theta_0) \geqslant k_1 p_1'(t, \theta_0) + k_2 p_1(t, \theta_0)$$

within w_0, and conversely outside it. The equality must hold on the boundary, which is equivalent to the theorem.

27.8. Writing

$$\phi = \left[\frac{\partial}{\partial \theta} \log p\right]_{\theta=\theta_0} \qquad (27.23)$$

$$\phi' = \left[\frac{\partial^2}{\partial \theta^2} \log p\right]_{\theta=\theta_0} \qquad (27.24)$$

we have

$$p'(\theta_0) = \phi\, p(\theta_0)$$
$$p''(\theta_0) = (\phi' + \phi^2)\, p(\theta_0)$$

and hence the inequality (27.15) reduces to

$$\phi' + \phi^2 \geqslant k_1 \phi + k_2 \quad . \quad . \quad . \quad . \quad . \quad (27.25)$$

within w_0, wherever $p(\theta_0)$ does not vanish; and conversely outside w_0.

We may distinguish three special cases :—

(*a*) If ϕ' is a function of ϕ, say $F(\phi)$, we have—

$$F(\phi) + \phi^2 \geqslant k_1 \phi + k_2, \quad . \quad . \quad . \quad . \quad . \quad (27.26)$$

and the Type A region is bounded by the surfaces

$$\phi_j = c_j \quad \text{and} \quad j = 1 \ldots m, \quad . \quad . \quad . \quad . \quad (27.27)$$

where m is the number of roots of (27.26). In this case, as we saw in **17.30**, there exists a sufficient estimator. It follows that w_0 is defined by inequalities of the type

$$c_1 \leqslant \phi \leqslant c_2,$$

and we may, as in **26.24**, use the ϕ's as new co-ordinates and calculate the size of a region from their distribution functions.

(*b*) As a simple case of (*a*), if

$$\phi' = A + B\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad (27.28)$$

we find, for (27.26),

$$\phi^2 - k_3 \phi - k_4 = 0, \quad . \quad . \quad . \quad . \quad . \quad (27.29)$$

and the limits of ϕ are given by the two roots of this quadratic.

(*c*) If ϕ' cannot be expressed as a function of ϕ which does not involve the x's explicitly, we shall have

$$\phi' \geqslant k_2 + k_1 \phi - \phi^2. \quad . \quad . \quad . \quad . \quad . \quad (27.30)$$

In this case, considering ϕ and ϕ' as two co-ordinates of a point in a plane, we see that the region for which (27.30) is true is the one "above" the parabola $\phi' = k_2 + k_1 \phi - \phi^2$, and that k_1, k_2 are determined by

$$\int_{-\infty}^{\infty} d\phi \int_{\phi'}^{\infty} p(\phi, \phi')\, d\phi' = 1 - \alpha \quad . \quad . \quad . \quad . \quad (27.31)$$

$$\int_{-\infty}^{\infty} \phi\, d\phi \int_{\phi'}^{\infty} p(\phi, \phi')\, d\phi' = 0. \quad . \quad . \quad . \quad . \quad (27.32)$$

In this instance we can reduce the problem to two dimensions by using two new co-ordinates ϕ, ϕ'.

Example 27.1

Consider the normal distribution

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}(x - \mu)^2\right\} dx.$$

To apply the foregoing theory with complete rigour we have to show that (27.14) is true. We shall assume that this is so, referring the reader for a formal proof to Neyman and Pearson (1936).

We have, then, with $\theta = \mu$,

$$\log p(\mu) = -\tfrac{1}{2} n \log(2\pi) - \tfrac{1}{2} \Sigma (x - \mu)^2$$
$$\phi = \Sigma (x - \mu_0), \qquad \phi' = -n,$$

and hence this case reduces to that of (27.28). We write

$$\phi = n(\bar{x} - \mu_0),$$

and can clearly use $\bar{x}$ instead of ϕ as a co-ordinate, which confirms the result of **27.7** since $\bar{x}$ is sufficient for μ.

It follows that the unbiassed region of Type A is given by

$$\bar{x} \leqslant \bar{x}_1, \qquad \bar{x} \geqslant \bar{x}_2$$

where
$$\int_{\bar{x}_1}^{\bar{x}_2} p(\bar{x})\, d\bar{x} = \alpha$$

and
$$\int_{\bar{x}_1}^{\bar{x}_2} p(\bar{x})(\bar{x} - \mu)\, d\bar{x} = 0.$$

Now if H_0 is true, that is if $\mu = \mu_0$, $\bar{x}$ is distributed in the form

$$dF = \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\bar{x} - \mu_0)^2\right\}.$$

Hence $\bar{x}_1 = -\bar{x}_2$ and the Type A region is defined as being *outside* the range

$$\mu_0 - \frac{\lambda}{\sqrt{n}} \leqslant \bar{x} \leqslant \mu_0 + \frac{\lambda}{\sqrt{n}}$$

where λ is given by

$$\int_\lambda^\infty \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx = \tfrac{1}{2}(1 - \alpha).$$

In this case the Type A test leads to the usual test based on equal tail areas. The same test follows from the likelihood ratio, as the reader can verify for himself.

Example 27.2

If the distribution is normal with zero mean and variance σ^2, and H_0 is that $\sigma = \sigma_0$, we find

$$\phi = \frac{n}{\sigma_0^3}\left\{\frac{1}{n}\Sigma(x^2) - \sigma_0^2\right\} = \frac{1}{\sigma_0}(v - n), \text{ say.}$$

This also satisfies (27.28), and the Type A region will be defined by

$$v_2 \leqslant v = \frac{1}{\sigma_0^2}\Sigma x^2, \quad \text{or } v \leqslant v_1,$$

where
$$\int_{v_1}^{v_2} p(v)\, dv = \alpha$$

and
$$\int_{v_1}^{v_2} p(v)(v - n)\, dv = 0.$$

Here $p(v)$, the frequency function of the second moment, is

$$p(v) = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} v^{\frac{1}{2}(n-2)} e^{-\frac{1}{2}v}\, dv,$$

and we find, for the second equation,

$$\int_{v_1}^{v_2} v^{\frac{1}{2}n} e^{-\frac{1}{2}v}\, dv - n\int_{v_1}^{v_2} v^{\frac{1}{2}n-1} e^{-\frac{1}{2}v}\, dv = 0.$$

Integrating the first member by parts, v being one part, we are left with

$$\left[-2v^{\frac{1}{2}n} e^{-\frac{1}{2}v}\right]_{v_1}^{v_2} = 0$$

or
$$v_1^{\frac{1}{2}n} e^{-\frac{1}{2}v_1} = v_2^{\frac{1}{2}n} e^{-\frac{1}{2}v_2}.$$

This has to be solved in conjunction with

$$\int_{v_1}^{v_2} \frac{1}{2^{\frac{1}{2}n}\,\Gamma\left(\frac{1}{2}n\right)}\, v^{\frac{1}{2}(n-2)}\, e^{-\frac{1}{2}v}\, dv = \alpha.$$

The numerical solution can be carried out by successive approximation or graphically.

In this connection Fig. 27.2 is of interest. It shows, for samples of two and $\alpha = 0{\cdot}98$, the graphs of the power function for the ordinary test with equal tail areas, in addition to the power functions for the Type A test, the U.M.P. test with $\sigma > \sigma_0$ and the U.M.P. test with $\sigma < \sigma_0$.

Evidently, for $\sigma > \sigma_0$ the best critical region (2) has the greatest power (as it must have), and for $\sigma < \sigma_0$ the best region (1) has the greatest power. The test based on equal

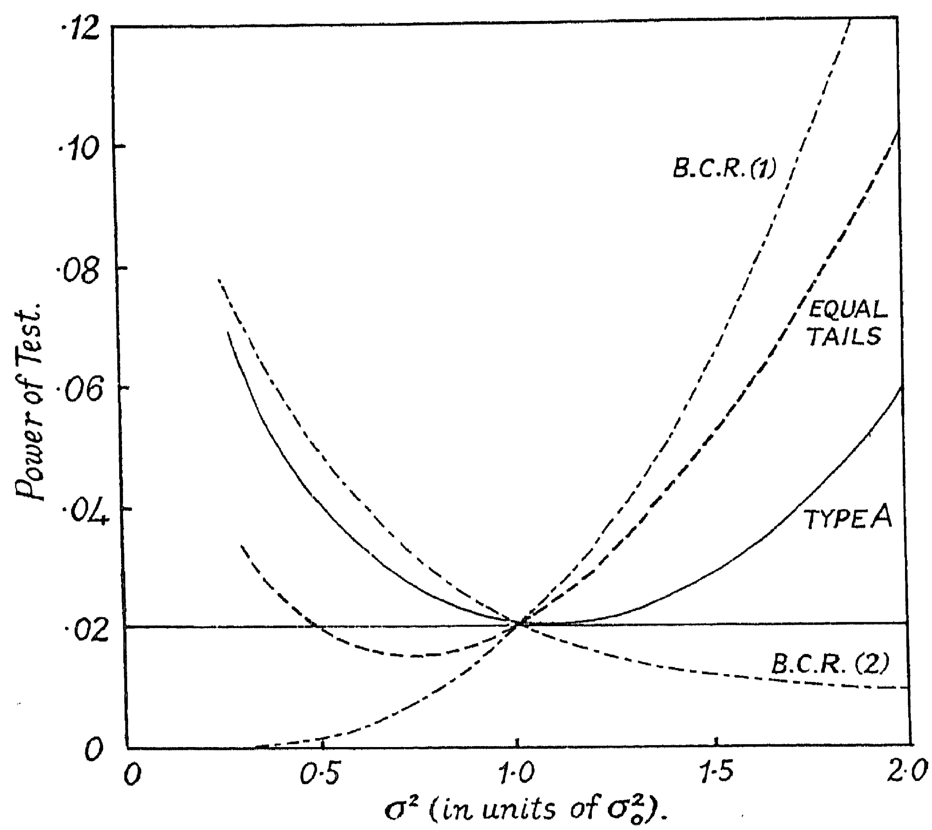


FIG. 27.2.—Power Curves of Four Different Tests of the Variance in Normal Samples of 2 (see text).

tail areas has a greater power than the Type A test for $\sigma > \sigma_0$ but a lower power for $\sigma < \sigma_0$, besides being biassed, as we have seen.

As n becomes larger the same effects persist, but the Type A and the "equal tails" tests become closer together in power. For samples of 20 or more there seems to be no serious loss in using the latter since the range of bias and its magnitude are then very small. If, of course, we knew in practice that $\sigma > \sigma_0$ we should use the U.M.P. test, and cases may arise, even when such knowledge is lacking, where "one-sided" hypotheses of this kind are all that concern us.

Invariance Theorem for Type A Regions

27.9. It is important to show that the regions selected on the basis of Type A criteria conform to corresponding criteria if some other function $\zeta(\theta)$ is used instead of θ itself. In Example 27.2, for instance, where we took θ to be the standard deviation σ, should we

have obtained the same regions if we had taken θ to be the variance σ^2? The answer is affirmative under certain general conditions, as we should expect from the relationship with sufficient estimators.

Suppose we have a new parameter ζ, given by

$$\theta = \theta_0 + f(\zeta) = \psi(\zeta), \qquad (27.33)$$

where $f(0) = 0$. Then if $p(\psi)$ satisfies (27.14) and the similar equation in second differentials, if ψ is monotonically increasing and $\left[\frac{\partial\psi}{\partial\zeta}\right]_0 > 0$, then the region based on ζ is an unbiassed critical region if that based on θ is so. It is sufficient to show that (27.15) and (27.16) are satisfied for ζ. Now

$$\theta = \psi(\zeta), \qquad \psi(0) = \theta_0, \qquad \left[\frac{\partial\psi}{\partial\zeta}\right]_0 = \psi' \neq 0, \qquad \left[\frac{\partial^2\psi}{\partial\zeta^2}\right]_0 = \psi'' \text{ (say)}. \qquad (27.34)$$

Thus

$$p'_\zeta(E \mid \theta_0) = p'_\zeta(E \mid \psi(0))$$
$$= p'_\theta(E \mid \theta_0)\,\psi',$$

and

$$p''_\zeta(E \mid \psi(0)) = p''_\theta(E \mid \theta_0)\,\psi'^2 + p'_\theta(E \mid \theta_0)\,\psi''.$$

Solving these for p'_θ and p''_θ and substituting in (27.15) and (27.16), we find

$$p''_\zeta(E \mid \psi(0)) \geqslant k_1'\, p'_\zeta(E \mid \psi(0)) + k_2'\, p_\zeta(E \mid \psi(0)) \qquad (27.35)$$

within w and the contrary outside, where

$$k_1' = \frac{k_1\,\psi'^2 + \psi''}{\psi'}, \qquad k_2' = k_2\,\psi'^2. \qquad (27.36)$$

The result follows.

Regions of Type A_1

27.10. The regions of Type A are determined so that tests based on them are U.M.P.U. in the neighbourhood of θ_0. We now consider a region, said to be of Type A_1, which is U.M.P.U. everywhere, i.e. which obeys (27.11) and (27.12) but has, in place of (27.13),

$$\int_{w_0} p\,dx \geqslant \int_{w} p\,dx \qquad (27.37)$$

for every admissible θ and every w satisfying the other two conditions.

It is conceivable that (27.37) does not entail the existence of a U.M.P.U. test, for there might be an unbiassed region of size $1-\alpha$ for which the derivative of $\int p\,dx$ did not exist at $\theta = \theta_0$ but which nevertheless gave a more powerful test. This refinement, however, need not detain us.

27.11. If W_+ represents the sample-space where the density is not zero, if

$$\phi' = A + B\phi,$$

and if $\phi(\theta_0)$ does not vanish identically in W_+ then the unbiassed critical region of Type A is necessarily of Type A_1.

Let w_0 be the Type A region, which is determined *ex hypothesi* by two numbers c_1 and c_2, such that—

$$c_1 \leqslant \phi_0 \leqslant c_2 \qquad \text{outside } w_0.$$

We have to show that

$$\int_{w_0} p\,dx \geqslant \int_{w} p\,dx$$

for all admissible θ and any w for which

$$\int_{w} p\,dx = 1 - \alpha, \qquad (27.38)$$

with the consequence that

$$\int_{w} p'\,dx = 0. \qquad (27.39)$$

Since $\phi' = A + B\phi$ we have, solving this equation as a linear differential equation of the first degree,

$$\phi = \left\{ \int A \exp\left(- \int B\,d\theta \right) d\theta + T \right\} \exp \int B\,d\theta. \qquad (27.40)$$

The reader may verify that this *is* a solution, and since it contains the arbitrary constant T it is the most general solution. It follows that we may write

$$\log p = P(\theta) + TQ(\theta) + f(x), \text{ say}, \qquad (27.41)$$

where P and Q do not depend upon x. We then have—primes denoting differentiation with respect to θ and the suffix 0 relating to θ_0—

$$\phi_0 = P_0' + TQ_0'. \qquad (27.42)$$

We note that Q_0' cannot be zero, for if it were we should have

$$0 = \int \phi_0 p_0\,dx = P_0' \int p_0\,dx = P_0',$$

which would imply that ϕ_0 was identically zero.

In virtue of the lemma of **27.6**, the proposition will be proved if we can show that for fixed θ and θ_0 there are two numbers a and b, depending on θ and θ_0 but not on the x's, such that

$$p \geqslant p_0 (a\phi_0 + b) \qquad \text{inside } w_0 \qquad (27.43)$$

and the contrary outside w_0. Putting the values of p and ϕ_0 in this expression, we have to show that a and b can be found such that, inside w_0,

$$\exp\{ P(\theta) + TQ(\theta) + f(x) \} \geqslant \exp\{ P(\theta_0) + TQ(\theta_0) + f(x) \} \{aP_0' + aTQ_0' + b\}$$

or, writing $r = P(\theta) - P(\theta_0)$, $q = Q(\theta) - Q(\theta_0)$, such that

$$\begin{aligned} \exp(r + qT) &\geqslant aQ_0'T + aP_0' + b \\ &\geqslant a_1 T + b_1, \text{ say}. \end{aligned} \qquad (27.44)$$

Here q cannot be zero, for if it were $Q(\theta)$ would be equal to $Q(\theta_0)$ and, integrating the frequency functions over W, we should find $r = 0$. The alternative hypothesis would not then differ essentially from H_0.

Consider at the outset the case when c_1 and c_2 are different. From (27.42) we see that ϕ_0 depends only on T so far as variation in x is concerned, and that

$$\text{if} \quad \phi_0 = c_1 \qquad T = \frac{c_1 - P_0'}{Q_0'} = T_1 \text{ (say)} \qquad (27.45)$$

$$\text{if} \quad \phi_0 = c_2 \qquad T = \frac{c_2 - P_0'}{Q_0'} = T_2 \text{ (say)}. \qquad (27.46)$$

T_1 and T_2 are different. Choose a_1 and b_1 so as to satisfy

$$\left.\begin{aligned} a_1 T_1 + b_1 &= e^{r+qT_1} \\ a_1 T_2 + b_1 &= e^{r+qT_2} \end{aligned}\right\} \qquad (27.47)$$

Then (27.44) is satisfied at the boundary points and we have merely to prove that

$$\left.\begin{aligned} c_1 < \phi_0 < c_2 \text{ implies } e^{r+qT} < a_1 T + b_1 \\ \phi_0 < c_1 \text{ and } \phi_0 > c_2 \text{ imply } e^{r+qT} > a_1 T + b_1 \end{aligned}\right\} \qquad (27.48)$$

This follows from the fact that

$$y = e^{r+qT} - a_1 T - b_1$$

has only one minimum, between T_1 and T_2, as may be seen by differentiating it twice, for the second derivative is positive and hence the first is a monotonically increasing function. But y vanishes at T_1 and T_2 and hence is negative between those values and positive outside them.

Finally, if c_1 and c_2 are equal, say to c, we choose a_1 and b_1 so as to satisfy

$$\left.\begin{aligned} P_0' + Q_0' T_0 &= c \\ q\, e^{r+qT_0} - a_1 &= 0 \\ e^{r+qT_0} - a_1 T_0 - b_1 &= 0 \end{aligned}\right\} \qquad (27.49)$$

It will be found that y has a minimum at $T = T_0$ and vanishes there. It follows that in the region ϖ_0 complementary to w_0, where $\theta_0 = c$, we have

$$e^{r+qT} = a_1 T + b_1,$$

and thus in w_0 where $\phi_0 \leqslant c$ or $c \leqslant \phi_0$ the left-hand side must be less than the right-hand side. The demonstration is complete.

Example 27.3

Consider again the data of Example 27.2. We have already seen that for this distribution $\phi' = A\phi + B$, so that the regions of Type A are also of Type A_1. Among unbiassed tests of the hypothesis this is the uniformly most powerful test.

Composite Hypotheses: Regions of Type B

27.12. We now consider the extension of the foregoing results to the case when H_0 is composite. For simplicity we will suppose that there are two parameters θ_1 and θ_2, H_0 specifying θ_1 as say θ_{10} and leaving θ_2 undetermined. Then a region w_0 will be said to be of Type B if

(*a*) $\int_{w_0} p(\theta_{10}, \theta_2)\, dx = 1 - \alpha$ for all admissible θ_2; . . . . . . (27.50)

(*b*) $\int_{w_0} p(\theta_1, \theta_2)\, dx$ may be differentiated twice with respect to θ_1 under the integral sign;

(*c*) $$\left[\frac{\partial}{\partial\theta_1}\int_{w_0} p(\theta_1, \theta_2)\, dx\right]_{\theta_1=\theta_{10}} = 0. \qquad (27.51)$$

(*d*) For any other region w satisfying (27.50),

$$\left[\frac{\partial^2}{\partial\theta_1^2}\int_{w_0} p\, dx\right]_{\theta_1=\theta_{10}} \geqslant \left[\frac{\partial^2}{\partial\theta_1^2}\int_{w} p\, dx\right]_{\theta_1=\theta_{10}} \qquad (27.52)$$

These conditions are obvious generalisations of those defining Type A. Putting now

$$\phi_j = \frac{\partial}{\partial \theta_j} \log p \qquad j = 1,\ 2 \quad . \quad . \quad . \quad . \quad (27.53)$$

$$\phi_{jk} = \frac{\partial \phi_j}{\partial \theta_k} = \phi_{kj}, \qquad k = 1,\ 2, \quad . \quad . \quad . \quad . \quad (27.54)$$

we state that the Type B region will exist and may be found if ϕ_1 and ϕ_2 are algebraically independent, if

$$\left.\begin{aligned} \phi_{11} &= A_0 + A_1\,\phi_1 + A_2\,\phi_2 \\ \phi_{12} &= B_0 + B_1\,\phi_1 + B_2\,\phi_2 \\ \phi_{22} &= C_0 \qquad\quad\;\; + C_2\,\phi_2 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (27.55)$$

and if the law of distribution of ϕ_2 is uniquely determined by its moments. We omit the proof of this theorem, for which see Neyman (1935*b*).

Simple Hypotheses with Two Parameters: Regions of Type C

27.13. The extension of the foregoing theory to the case of a simple hypothesis specifying several parameters presents some new features. Again to simplify the discussion we shall consider two parameters, θ_1 and θ_2.

Consider the power function in the neighbourhood of $\theta_1 = \theta_2 = 0$ which we will suppose to be the values specified by H_0. Writing for the function

$$\beta\,(\theta_1,\ \theta_2 \mid w) = \int_w p\,(\theta_1,\ \theta_2)\,dx \quad . \quad . \quad . \quad . \quad . \quad (27.56)$$

$$\left[\frac{\partial \beta}{\partial \theta_j}\right]_{\theta_1=\theta_2=0} = \beta_j, \qquad j = 1,\ 2 \quad . \quad . \quad . \quad . \quad (27.57)$$

$$\left[\frac{\partial^2 \beta}{\partial \theta_j\,\partial \theta_k}\right]_{\theta_1=\theta_2=0} = \beta_{jk}, \qquad j,\ k = 1,\ 2 \quad . \quad . \quad . \quad (27.58)$$

we have, assuming an expansion by Taylor's theorem,

$$\begin{aligned}\beta\,(\theta_1,\ \theta_2 \mid w) &= \beta\,(0,\ 0 \mid w) + \theta_1\,\beta_1\,(w) + \theta_2\,\beta_2\,(w) \\ &\quad + \tfrac{1}{2}\,\{\theta_1^2\,\beta_{11}\,(w) + 2\theta_1\,\theta_2\,\beta_{12}\,(w) + \theta_2^2\,\beta_{22}\,(w)\,\} + \ldots \quad . \quad . \quad (27.59)\end{aligned}$$

To extend the idea of unbiassed tests to such a case we require in the first place

$$\left.\begin{aligned}\beta_1\,(w) &= 0 \\ \beta_2\,(w) &= 0\end{aligned}\right\}. \quad . \quad . \quad . \quad . \quad . \quad (27.60)$$

Secondly, there will be a minimum at $\theta_1 = \theta_2 = 0$ if

$$\Delta = \beta_{12}^2 - \beta_{11}\,\beta_{22} < 0 \quad . \quad . \quad . \quad . \quad . \quad (27.61)$$

and

$$\beta_{11},\ \beta_{22} > 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (27.62)$$

If these conditions are satisfied the power function for small values of θ_1 and θ_2 is effectively

$$\beta\,(\theta_1,\ \theta_2 \mid w) = 1 - \alpha + \tfrac{1}{2}\,\{\theta_1^2\,\beta_{11} + 2\theta_1\,\theta_2\,\beta_{12} + \theta_2^2\,\beta_{22}\} \quad . \quad . \quad (27.63)$$

We may represent this diagrammatically as in Fig. 27.3, which shows one of the ellipses for which the power function is constant.

Since the hypothesis H_0 is that $\theta_1 = \theta_2 = 0$, we may speak of the value θ_1 as the "error in θ_1", and similarly for θ_2; and if, as in the case depicted, the co-ordinate axes are not the same as the principal axes of the ellipse it is clear that for values of θ_1 which are not zero, errors of positive and negative sign in θ_2 are not equal. From this viewpoint it may

be said that the minimisation of the power function does not control positive or negative errors to the same extent ; for the points A and B in Fig. 27.3 lie on the ellipse of constant

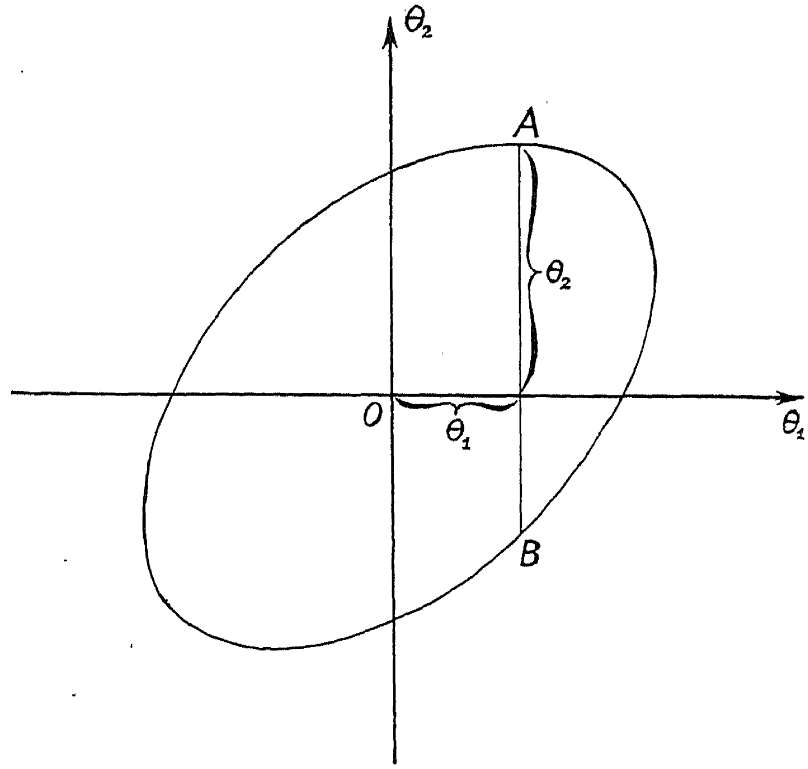


FIG. 27.3.—Ellipse of Constant Power for Simple Hypothesis with Two Parameters (see text).

β, so that the probability of detecting them is the same, though A represents a positive " error " in θ_2 greater than the negative " error " given by B.

27.14. Whether this is a desirable property of the test depends to some extent on what the test is intended to do. To avoid the anomaly we must require that

$$\beta_{12} = 0. \qquad (27.64)$$

Furthermore, even if this condition is satisfied and the principal axes of the ellipse coincide with the co-ordinate axes, there may still appear anomalies if the length of one axis is greater than that of the other ; for then errors in one parameter are not detected as frequently as errors *of the same size* in the other. Here again it is a matter of particular circumstance whether such an effect is regarded as objectionable. (We disregard the fact that it can be removed by appropriate scaling of the parameters, which may or may not be artificial.) To remove it we must require that

$$\beta_{11} = \beta_{22}, \qquad (27.65)$$

so that the ellipses reduce to circles.

We may refer to the ellipses as " curves of equidetectability."

27.15. With the foregoing explanation in mind we define w_0 as a *regular* unbiassed critical region of Type C if it obeys the conditions

$$\beta_1 (w_0) = \beta_2 (w_0) = 0 \qquad (27.66)$$

$$\beta_{12} (w_0) = 0 \qquad (27.67)$$

$$\beta_{11} (w_0) = \beta_{22} (w_0) \qquad (27.68)$$

and if, for any other region obeying these three conditions and for which

$$\beta(0, 0 \mid w_0) = \beta(0, 0 \mid w) = 1 - \alpha, \quad . \quad . \quad . \quad . \quad (27.69)$$

we have

$$\beta_{11}(w_0) \geqslant \beta_{11}(w). \quad . \quad . \quad . \quad . \quad . \quad . \quad (27.70)$$

Secondly, if a region w_1 possesses the property that

$$\beta_1(w_1) = \beta_2(w_1) = 0 \quad . \quad . \quad . \quad . \quad . \quad (27.71)$$

$$\beta_{12}^2(w_1) - \beta_{11}(w_1)\,\beta_{22}(w_1) < 0 \quad . \quad . \quad . \quad . \quad . \quad (27.72)$$

and for any other region obeying the conditions

$$\beta(0, 0 \mid w_1) = \beta(0, 0 \mid w) = 1 - \alpha \quad . \quad . \quad . \quad . \quad (27.73)$$

$$\frac{\beta_{11}(w_1)}{\beta_{11}(w)} = \frac{\beta_{12}(w_1)}{\beta_{12}(w)} = \frac{\beta_{22}(w_1)}{\beta_{22}(w)} \quad . \quad . \quad . \quad . \quad . \quad (27.74)$$

we have

$$\beta_{11}(w_1) \geqslant \beta_{11}(w) \quad . \quad . \quad . \quad . \quad . \quad . \quad (27.75)$$

we shall say that w_1 is a *non-regular* unbiassed critical region of Type C.

These equations are analytical ways of saying that the regular region of Type C is the one, among all regions having circular curves of equidetectability, which has the smallest radius for any given value of the power function ; whereas the non-regular region of Type C is the one, among all regions having similar ellipses of equidetectability, which has the smallest axes.

27.16. We now state without proof theorems similar to those demonstrated above for the case of a single parameter.

Write

$$p_{jk} = \left[\frac{\partial^2 p}{\partial\theta_j\,\partial\theta_k}\right]_{\theta_1=\theta_2=0} \quad \text{etc.}$$

Then w_0 is a regular unbiassed critical region of Type C if

(*a*) inside w_0

$$p_{11} \geqslant k_1(p_{11} - p_{22}) + k_2\,p_{12} + k_3\,p_1 + k_4\,p_2 + k_5\,p, \quad . \quad . \quad (27.76)$$

and outside w_0 the inequality is reversed—

$$(b) \qquad \int_{w_0} p_j\,dx = \int_{w_0} p_{12}\,dx = \int_{w_0} (p_{11} - p_{22})\,dx = 0, \qquad j = 1, 2, \qquad (27.77)$$

Secondly, if w_1 satisfies the conditions—

(*a*) that inside w_1

$$p_{11} \geqslant k_1(\gamma_{12}\,p_{11} - \gamma_{11}\,p_{12}) + k_2(\gamma_{22}\,p_{11} - \gamma_{11}\,p_{22}) + k_3\,p_1 + k_4\,p_2 + k_5\,p \quad (27.78)$$

and outside w_1 the inequality is reversed, the k's as usual being constants and the γ's obeying the conditions

$$\gamma_{11} > 0, \qquad \gamma_{12}^2 - \gamma_{11}\,\gamma_{22} < 0\,;$$

$$(b) \qquad \int_{w_1} p_j\,dx = \int_{w_1} (\gamma_{12}\,p_{11} - \gamma_{11}\,p_{12})\,dx = \int_{w_1} (\gamma_{22}\,p_{11} - \gamma_{11}\,p_{22})\,dx = 0, \qquad (27.79)$$

then w_1 is a non-regular unbiassed critical region of Type C, having ellipses of equidetectability determined by

$$\gamma_{11}\,\theta_1^2 + 2\gamma_{12}\,\theta_1\,\theta_2 + \gamma_{22}^2\,\theta_2^2 = \text{constant}. \quad . \quad . \quad . \quad (27.80)$$

27.17. The theorem of invariance of **27.9** no longer holds in general for the present case. If we transform to new parameters ζ_1 and ζ_2, the equations of transformation

$$d\zeta_1 = \frac{\partial \zeta_1}{\partial \theta_1} d\theta_1 + \frac{\partial \zeta_1}{\partial \theta_2} d\theta_2,$$

etc. will not transform an ellipse co-axial with the co-ordinate axes θ_1, θ_2 into one co-axial with ζ_1, ζ_2. Thus, in general, the effect of a transformation is to make a regular Type C region into a non-regular Type C region.

27.18. As usual, the conditions for the Type C region may be simply written in terms of the derivatives of $\log p$. Write

$$\phi_j = \left[\frac{\partial}{\partial \theta_j} \log p\right]_{\theta_1=\theta_2=0} \quad . \quad . \quad . \quad . \quad . \quad (27.81)$$

$$\phi_{jk} = \left[\frac{\partial^2 \log p}{\partial \theta_j \, \partial \theta_k}\right]_{\theta_1=\theta_2=0} \quad . \quad . \quad . \quad . \quad . \quad (27.82)$$

Then if

$$\phi_{jk} = A_{jk} + B_{jk}\,\phi_1 + C_{jk}\,\phi_2 \quad . \quad . \quad . \quad . \quad . \quad (27.83)$$

we shall have

$$p_{jk} = (\phi_j\,\phi_k + A_{jk} + B_{jk}\,\phi_1 + C_{jk}\,\phi_2)\,p \quad . \quad . \quad . \quad (27.84)$$

and the inequality (27.76) becomes

$$(1 - k_1)\,\phi_1^2 - k_2\,\phi_1\,\phi_2 + k_1\,\phi_2^2 - k_3'\,\phi_1 - k_4'\,\phi_2 - k_5' \geqslant 0 \quad . \quad . \quad (27.85)$$

where the k' are new constants easily expressible in terms of the old. They must be determined so as to satisfy (27.77), which reduce to

$$\int_{w_0} \phi_j\, p\, dx = \int_{w_0} (\phi_1\,\phi_2 + A_{12})\, p\, dx = \int_{w_0} \{\phi_1^2 - \phi_2^2 + (A_{11} - A_{22})\}\, p\, dx = 0. \quad (27.86)$$

Example 27.4

Suppose we have a sample of n_1 from a normal population with mean μ_1 and unit variance and a second sample of n_2 from a normal population with mean μ_2 and also unit variance. The simple hypothesis to be tested is $\mu_1 = \mu_2 = \mu_0$, where μ_0 is some specified value. We consider two cases :—

(i) in which errors of the same size in μ_1 and μ_2 are equally important ;

(ii) in which, for some reason, there is a stronger desire to avoid errors in μ_2 than in μ_1 and that therefore a greater number n_2 of members has been taken in the second sample. We also assume that the sizes of errors judged of equal importance are inversely proportional to $\sqrt{n}$, so that we are led to consider new parameters—

$$\eta_1 = (\mu_1 - \mu_0)\sqrt{n_1}, \qquad \eta_2 = (\mu_2 - \mu_0)\sqrt{n_2} \quad . \quad . \quad . \quad (27.87)$$

Case 1.—The frequency function is

$$p \propto \exp\left[-\tfrac{1}{2}\sum_{1}^{n_1}(x_j - \mu_1)^2 - \tfrac{1}{2}\sum_{n_1+1}^{n_1+n_2}(x_j - \mu_2)^2\right].$$

It will be found that

$$\phi_1 = n_1(\bar{x}_1 - \mu_0)\,; \qquad \phi_2 = n_2(\bar{x}_2 - \mu_0)\,;$$

$$\phi_{11} = -n_1 = A_{11}, \qquad \phi_{12} = 0 = A_{12}\,; \qquad \phi_{22} = -n_2 = A_{22}.$$

From (27.85) we then find

$$(1 - k_1)\, n_1^2\, (\bar{x}_1 - \mu_0)^2 - k_2\, n_1\, n_2\, (\bar{x}_1 - \mu_0)\, (\bar{x}_2 - \mu_0) + k_1\, n_2^2\, (\bar{x}_2 - \mu_0)^2 - k_3'\, n_1\, (\bar{x}_1 - \mu_0) - k_4'\, n_2\, (\bar{x}_2 - \mu_0) - k_5' \geqslant 0. \quad (27.88)$$

The law of distribution of $\bar{x}_1$ and $\bar{x}_2$ may be written

$$p \propto \exp\left[-\tfrac{1}{2}\, \{n_1\, (\bar{x}_1 - \mu_0)^2 + n_2\, (\bar{x}_2 - \mu_0)^2\}\right]. \quad (27.89)$$

Put $\qquad u = \sqrt{n_1}\, (\bar{x}_1 - \mu_0)$ and $v = \sqrt{n_2}\, (\bar{x}_2 - \mu_0)$.

Then the region w_0 is determined by

$$(1 - k_1)\, n_1\, u^2 - k_2\, uv\, \sqrt{(n_1\, n_2)} + k_1\, n_2\, v^2 - k_3'\, u\sqrt{n_1} - k_4'\, v\sqrt{n_2} - k_5' \geqslant 0 \quad (27.90)$$

where

$$\int_{w_0} p\, (u, v)\, du\, dv = 1 - \alpha$$

$$\int_{w_0} u\, p\, (u, v)\, du\, dv = \int_{w_0} v\, p\, (u, v)\, du\, dv = \int_{w_0} uv\, p\, (u, v)\, du\, dv = 0 \quad (27.91)$$

$$\int_{w_0} (n_1\, u^2 - n_2\, v^2)\, p\, (u, v)\, du\, dv = (1 - \alpha)\, (n_1 - n_2) \quad (27.92)$$

and

$$p\, (u, v) = \frac{1}{2\pi} \exp\{-\tfrac{1}{2}\, (u^2 + v^2)\}.$$

It is evident from (27.90) that in the (u, v) plane the boundary of w_0 is a conic. From (27.91) we see that it must be coaxial with the co-ordinate axes and have its centre at the origin. Hence $k_2 = k_3' = k_4' = 0$. Finally from (27.92) we find that the boundary is of the form

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1, \quad (27.93)$$

where

$$\frac{1}{a^2} = \frac{n_1\, (1 - k_1)}{k_5'}, \qquad \frac{1}{b^2} = \frac{n_2\, k_1}{k_5'}. \quad (27.94)$$

The Type C regions are then defined by (27.93), but we have to express a and b in terms of known constants, including the probability level $1 - \alpha$. We have to satisfy (27.92), and will show that a solution always exists.

Put

$$F\, (a, b) = \frac{1}{2\pi} \int_{w_0} (n_1\, u^2 - n_2\, v^2) \exp\{-\tfrac{1}{2}\, (u^2 + v^2)\}\, du\, dv - (n_1 - n_2)\, (1 - \alpha). \quad (27.95)$$

If the boundary of w_0 is a circle, its radius is easily found to be

$$a = b = \sqrt{\{-2 \log\, (1 - \alpha)\}}.$$

The integral $F\, (a, b)$ outside this circle, by the substitution $u = r \cos \psi$, $v = r \sin \psi$, is found to be

$$F\, (a, a) = (n_1 - n_2) \frac{1}{2\pi} \int_{u^2 + v^2 \geqslant a^2} u^2 \exp\{-\tfrac{1}{2}\, (u^2 + v^2)\}\, du\, dv - (n_1 - n_2)\, (1 - \alpha)$$

$$= (1 - \alpha)\, (n_1 - n_2)\, \tfrac{1}{2} a^2.$$

Now taking w_0 as the space outside the parallel lines

$$v = \pm \lambda,$$

which is given by a infinite, so that $\dfrac{2}{\sqrt{(2\pi)}} \displaystyle\int_\lambda^\infty e^{-\frac{1}{2}x^2}\, dx = 1 - \alpha,$

$$F(\infty, \lambda) = -(n_1 - n_2)(1-\alpha) + \frac{n_1}{2\pi}\int_{w_0} u^2 \exp\{-\tfrac{1}{2}(u^2+v^2)\}\,du\,dv$$
$$- \frac{n_2}{2\pi}\int_{w_0} v^2 \exp\{-\tfrac{1}{2}(u^2+v^2)\}\,du\,dv$$
$$= -n_2\sqrt{\frac{2}{\pi}}\,\lambda\, e^{-\frac{1}{2}\lambda^2} < 0.$$

Similarly,

$$F(\lambda, \infty) = n_1\sqrt{\frac{2}{\pi}}\,\lambda\, e^{-\frac{1}{2}\lambda^2} > 0.$$

Thus, since $F(a, b)$ is continuous it must vanish somewhere in the range $\lambda \leqslant a \leqslant \infty$, $\lambda \leqslant b \leqslant \infty$. The values for which it does so define the Type C region.

Case 2.—In this case, using the parameters η_1 and η_2 of (27.87), we find

$$\phi_1 = u, \qquad \phi_2 = v$$
$$\phi_{11} = -1, \qquad \phi_{12} = 0, \qquad \phi_{22} = -1.$$

The inequality becomes

$$(1-k_1)u^2 - k_2 uv + k_1 v^2 - k_3' u - k_4' v - k_5' \geqslant 0,$$

where

$$\int_{w_0} (u^2 - v^2)\,p(u, v)\,du\,dv = 0.$$

In a similar way it follows that the Type C region is the one lying outside the circle

$$u^2 + v^2 = -2\log(1-\alpha).$$

We leave the verification of this result to the reader.

Certain Limiting Properties

27.19. From the foregoing examples it will be seen that in certain cases the optimum critical regions are by no means easy to determine numerically; and it is not always clear that the labour involved is repaid by the results. Some consideration has been given by various writers to tests which have optimum properties for large n, the presumption being that the same tests will be good, if not the best, for small values. As usual when several limiting processes are involved simultaneously, the rigorous enunciation and proof of theorems in this field is a matter of some complexity, and we shall here merely indicate some of the results in very general terms without including proofs.

It has been shown by Neyman (1938*b*) that there do exist tests which are unbiassed in the limit, and rules have been given for finding them. It has also been shown by Wald (1941*a*) that there exist tests which are most powerful in the limit, and that such as are based on maximum likelihood estimators are of this class. The tests are uniformly most powerful for the single parameter $\theta > \theta_0$ and for $\theta < \theta_0$, but not both; and for any range they are the most powerful unbiassed tests in the limit. Furthermore, the Type A test tends to the most powerful unbiassed form.

The general conclusion seems to be that, even where the variation is not normal, most of the tests in current use which are based on likelihood estimators have optimum properties in the limit, and may therefore be used confidently for moderate or large samples. For small samples the position is not so clear, particularly for non-normal variation. Tests based on inefficient estimators are presumably less satisfactory; and for the non-parametric case there is as yet no complete theory. On this latter question reference may be made to a useful review by Scheffé (1943).

The Unbiassed Character of Likelihood-ratio Tests

27.20. It is of some interest to consider how far the tests based on likelihood (**26.35**) are unbiassed.

It has been shown (Pitman, 1939*b*; Brown, 1939) that the Neyman–Pearson test in the problem of k samples based on λ_{H_1} is biassed unless all the samples are of the same size; but that Bartlett's modification (**26.42**) is unbiassed. We prove this in **27.25** below. On the other hand, Daly (1940) has shown that in certain multivariate tests such as those of regressions, multiple correlations, Hotelling's T (which we introduce in the next chapter), and the ordinary analysis of variance and covariance for orthogonal or non-orthogonal data, the likelihood-ratio tests are unbiassed, at least in the Type A sense (i.e. locally) and in some cases completely so.

Pitman's Method for Location and Scale Parameters

27.21. In the special but not uncommon case where the hypotheses under test concern parameters of scale or location, a simplified approach is possible. Suppose the joint distribution of k sample-values is

$$dF = f(x_1 - \theta_1, x_2 - \theta_2, \ldots x_k - \theta_k)\, dx_1 \ldots dx_k. \qquad (27.96)$$

We seek for a statistic J, independent of the θ's, to test the hypothesis; and clearly, if the test is to be satisfactory, J must be independent of the origin, i.e. must be seminvariant. The test that the θ's are all equal is then equivalent to testing the hypothesis

$$\theta_1 = \theta_2 = \ldots = \theta_k = 0. \qquad (27.97)$$

Without loss of generality we may suppose the hypothesis rejected if J is small and less than some quantity depending on the acceptance value α, and we may also suppose J positive; for if either condition is not satisfied we can transfer to some other function of J for which it is.

In the sample space W, J must be constant along the line $x_1 = x_2 = \ldots = x_k =$ constant, and therefore the critical region w_0 will be the one lying outside a hypercylinder whose axis is parallel to this line. When H_0 is true, the probability of rejection is then

$$\int_{w_0} dF(x_1 \ldots x_k) = 1 - \alpha, \qquad (27.98)$$

and when it is not true the probability is

$$\int_{w_0} dF(x_1 - \theta_1, \ldots x_k - \theta_k)$$

$$= \int_{w} dF(x_1, \ldots x_k), \qquad (27.99)$$

where w is merely derived from w_0 by a translation in W without rotation. If L is any line parallel to $x_1 = \ldots = x_k = 0$, we write

$$P(L) = \int_L dF(x_1 \ldots x_k)$$

$$= \int_L f(x_1 \ldots x_k)\, d\eta \qquad (27.100)$$

where

$$\eta = \frac{1}{\sqrt{k}} \Sigma(x); \qquad (27.101)$$

and η is thus the distance of the point $(x_1 \ldots x_k)$ from the plane $\Sigma(x) = 0$.

Now if w_0 is defined as the locus of all lines for which $P(L) \geqslant h$, a constant, $P(L)$ will be less than h on any L which is in w but not in w_0. Hence

$$\int_{w_0} dF > \int_{w} dF, \qquad (27.102)$$

and so the resulting test is unbiassed. Thus an unbiassed test is given by choosing J so that at any point of a line L it is equal to $P(L)$ at that point. Now we may write for the variable co-ordinate on a particular L, say ξ_r,

$$\xi_r = x_r - t$$

where

$$t = \frac{1}{k}\Sigma(x) - \frac{\eta}{\sqrt{k}}.$$

Hence

$$P(L) = \sqrt{k}\int_{-\infty}^{\infty} f(x_1 - t, x_2 - t, \ldots x_k - t)\,dt. \qquad (27.103)$$

Taking

$$J = \frac{1}{\sqrt{k}}P(L),$$

we find

$$J = \int_{-\infty}^{\infty} f(x_1 - t, x_2 - t, \ldots x_k - t)\,dt, \qquad (27.104)$$

which gives us an unbiassed test.

Example 27.5

Consider the case where the variables are distributed normally with unit variance.

$$f = \frac{1}{(2\pi)^{\frac{k}{2}}}\exp\{-\tfrac{1}{2}\Sigma(x_j - \theta_j)^2\}.$$

Then we have, from (27.104),

$$J = \frac{1}{(2\pi)^{\frac{k}{2}}}\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}\Sigma(x_j - t)^2\}\,dt$$

$$= \frac{e^{-\frac{1}{2}S}}{\sqrt{k}\,(2\pi)^{\frac{1}{2}(k-1)}}$$

where

$$S = \Sigma(x - \bar{x})^2.$$

In practice we should take S as our criterion, not J, and reject the hypothesis that the means were unequal if S exceeded some fixed value determined by α. We observe that in fact S is distributed as χ^2 with $k - 1$ degrees of freedom when H_0 is true, so that this value is easily ascertained.

27.22. Consider now the case where the frequency function is

$$\frac{1}{\theta_1\,\theta_2 \ldots \theta_k} f\left(\frac{x_1}{\theta_1} \ldots \frac{x_k}{\theta_k}\right). \qquad (27.105)$$

If the x's are positive in range we put

$$y_j = \log x_j, \qquad \phi_j = \log\theta_j, \qquad (27.106)$$

and for the frequency function of the y's we find

$$\exp(\Sigma y - \Sigma\phi)\,f(e^{y_1-\phi_1}, e^{y_2-\phi_2}, \ldots e^{y_k-\phi_k}). \qquad (27.107)$$

This reduces to our first case, and we have an unbiassed criterion that

$$\phi_1 = \phi_2 = \ldots = \phi_k$$

by putting

$$J = \int_{-\infty}^{\infty} \exp(\Sigma y - kt) f(e^{y_1 - t}, e^{y_2 - t}, \ldots e^{y_k - t})\, dt$$

$$= \left(\prod_{j=1}^{k} x_j\right) \int_0^{\infty} f\left(\frac{x_1}{t}, \frac{x_2}{t}, \ldots \frac{x_k}{t}\right) \frac{dt}{t^{k+1}}. \quad . \quad . \quad . \quad (27.108)$$

When the x's are not necessarily positive the expression remains the same, except that in (27.108) $\Pi(x)$ becomes $\Pi(|x|)$. Small values of J are significant.

27.23. Suppose now that our hypothesis asserts the equality of θ's or ϕ's and states that they have a common value θ_0 or ϕ_0, as the case may be. Then if we take

$$J' = \left(\prod_{j=1}^{k} |x_j|\right) f(x_1 \ldots x_k), \quad . \quad . \quad . \quad . \quad (27.109)$$

the test will be unbiassed. Moreover, if we regard small values of J' as significant and the x's are independent, and if each frequency function is unimodal, then when

$$\theta_1 = \theta_2 = \ldots = \theta_k = \theta_0$$

is not true the probability that J' exceeds the specified limit based on $1 - \alpha$ increases as any θ tends to θ_0. J' therefore provides an unbiassed test.

27.24. Finally, consider the case of k variates each distributed in the form typified by

$$dF = \frac{1}{\phi_j \Gamma(m_j)} \exp\left(-\frac{x_j}{\phi_j}\right) \left(\frac{x_j}{\phi_j}\right)^{m_j - 1} dx_j. \quad . \quad . \quad . \quad (27.110)$$

Their joint distribution is

$$dF = \frac{\Pi\left(\frac{x}{\phi}\right)^{m-1} \exp\left(-\Sigma \frac{x}{\phi}\right) \Pi\, dx}{\Pi\{\phi \Gamma(m)\}}. \quad . \quad . \quad . \quad (27.111)$$

Hence, to test the hypothesis that the samples have the same ϕ we have

$$J = \frac{\Pi(x^m)}{\Pi\{\Gamma(m)\}} \int_0^{\infty} e^{-\Sigma(x)/t} \frac{dt}{t^{M+1}},$$

where $M = \Sigma(m)$,

$$= \frac{\Gamma(M)}{\Pi\{\Gamma(m)\}} \cdot \frac{\Pi(x^m)}{(\Sigma x)^M}. \quad . \quad . \quad . \quad . \quad . \quad (27.112)$$

It is sometimes convenient to deal with

$$K = \frac{\Pi(x^m)}{(\Sigma x)^M}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (27.113)$$

which differs from J only by a constant factor.

The maximum value of K is

$$\frac{\Pi(m^m)}{M^M}$$

and we put

$$L = -\frac{\log K}{\log \max. K} = M \log\left(\frac{\Sigma x}{M}\right) - \Sigma\left(m \log \frac{x}{m}\right) \quad . \quad . \quad (27.114)$$

L is essentially not negative, and large values are significant.

For testing the hypothesis that a set of variances have some *specified* equal value, we find similarly from (27.109)

$$L' = \Sigma (x) - M - \Sigma \left(m \log \frac{x}{m} \right). \qquad (27.115)$$

27.25. The foregoing result has an immediate application to the case of k normal samples, for the variances are then distributed in the Type III form of equation (27.110). The criterion L becomes

$$L = N \log \left(\frac{\Sigma (s_i^2)}{N} \right) - \Sigma \left(\nu_i \log \frac{s_i^2}{\nu_i} \right), \qquad (27.116)$$

where ν as usual represents the number of degrees of freedom and $N = \Sigma (\nu)$. This, as will be seen by comparison with (26.93), is equivalent to Bartlett's test, and shows that it is unbiassed.

NOTES AND REFERENCES

For the theory of unbiassed tests see particularly Neyman and Pearson (1936; 1938) and Neyman (1935*b*). Regions of Type B have also been considered by Scheffé (1942*a*), who discusses a Type B_1 standing in relation to B as Type A_1 to Type A.

For limiting properties see Neyman (1938*b*) and Wald (1941*a*).

See also references to the previous chapter.

EXERCISES

27.1. Show that the test of Example 27.1 provides regions which are of Type A_1 as well as of Type A, and that the test is a U.M.P.U. one.

27.2. Show that the cumulants of the distribution of L of (27.114) are

$$\kappa_1 = M \{G_1 (M) - \log M\} - \Sigma [m \{G_1 (m) - \log m\}]$$
$$\kappa_r = (-1)^r \{\Sigma m^r G_r (m) - M^r G_r (M)\}, \qquad r > 1$$

where
$$G_r = \frac{d^r}{dm^r} \log \Gamma (m).$$

Hence show that the cumulants of $\dfrac{L}{1+\beta}$ are approximately $\kappa_r = \dfrac{k-1}{2} \Gamma (r)$, where

$$\beta = \frac{1}{6(k-1)} \left\{ \Sigma \left(\frac{1}{m} \right) - \frac{1}{M} \right\},$$

and thus that $\dfrac{2L}{1+\beta}$ is distributed approximately as χ^2 with $k-1$ degrees of freedom.

(Bartlett, 1937*c*; Pitman, 1939*b*.)

27.3. Show that in samples of 3 from a normal population the distribution of the range r is given by—

$$dF = \frac{6}{\sigma\sqrt{\pi}} e^{-\frac{r^2}{4\sigma^2}} \int_0^{\frac{r}{\sigma\sqrt{6}}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y^2}\, dy\, dr.$$

Hence that an unbiassed critical region of Type **A** is given by

$$\left[r\, e^{-\frac{1}{4}r^2} \int_0^{\frac{r}{\sqrt{6}}} e^{-\frac{1}{2}y^2}\, dy \right]_{r_1}^{r_2} = 0$$

$$r_1\, e^{-\frac{1}{4}r_1^2} \int_0^{\frac{r_1}{\sqrt{6}}} e^{-\frac{1}{2}y^2}\, dy = r_2\, e^{-\frac{1}{4}r_2^2} \int_0^{\frac{r_2}{\sqrt{6}}} e^{-\frac{1}{2}y^2}\, dy,$$

the region lying outside $r_1 \leqslant r \leqslant r_2$.

(Neyman and Pearson, 1936.)

CHAPTER 28

MULTIVARIATE ANALYSIS

28.1. We have already considered some aspects of the case in which each member of a population is characterised by several variates $x_1 \ldots x_p$. For instance, we have examined the measurement of correlation between the variates and the regression of one variate on some or all of the others. In this chapter we shall extend our inquiries into the multivariate case a good deal further, mainly by taking into account the possibility that different sample-members may have emanated from different populations. This will lead to some generalisations of the methods already discussed for the univariate case, such as tests of homogeneity and tests of differences between two samples. Some of our known results generalise with nothing more than additional mathematical complexity; but in others certain new features appear, and the theory of multivariate analysis is not entirely a matter of generalising univariate results to p dimensions.

28.2. One or two examples will illustrate the kind of problem with which we are concerned. A number of skulls are discovered in a burial-ground. They are found to vary among themselves in the manner usual in biological material. Is the observed variation consistent with the hypothesis that all the skulls were derived from members of the same race or does it suggest a mixture of racial types? If heterogeneity is indicated, do the skulls fall into two well-defined categories, such as we might expect if the burial-ground were the site of a battle between two races such as Saxon and Celt; or are there several types such as we should expect in the normal burial-ground of a town where races were living together and interbreeding? Or again, if the skulls are compared with another set known to have been buried at a much earlier time from the same race, is there any evidence of a significant change in skulls from one period to the other?

There is no single measurement on a skull which is marked out from the infinite number of possible measurements for deciding questions of this kind. It is quite common for thirty or forty measurements to be taken by craniometricians on a single skull. Even if we reject many of these for practical reasons, leaving out the jawbone, for instance, because it is often separated from the skull and cannot be identified, we shall still be left with a number p which require consideration. For n skulls we shall then have n sets of p values corresponding to variates $x_1 \ldots x_p$ which are, in general, correlated among themselves and may be highly so. Our problem is to test the homogeneity of these values, or to estimate differences between parent populations from which they were derived. We may, of course, apply methods which are already familiar by picking out one variate and testing for homogeneity. But we might pick out quite an unsuitable one and sacrifice most of the information. Even if time permits we cannot take each variate in turn and test it because the variates are correlated and our p tests are not independent.

28.3. Again, suppose we have two different breeds of laying hen and are given a batch of eggs from the hen-run without knowing which hen laid which egg. We require to allocate the eggs to the two breeds. Assuming that there is no decisive criterion such as colour of shell, we may measure various properties of the eggs such as length, breadth,

weight, volume, specific gravity and so on. Some of these measurements will be highly correlated or, in the extreme case, perfectly correlated, as with weight, volume and specific gravity. In such circumstances we may reject some variates as redundant ; but in general we shall be left with several sets of measurements. Our problem is to find some method based on the retained variates for allocating the eggs to the correct parent breed. In particular we might search for the *best* linear function of the variates to discriminate between breeds and to enable us to assign the eggs with the maximum probability of correctness.

28.4. Throughout the whole chapter we shall, except when the contrary is stated, assume that the variation is normal. In addition, to render our formulae a little less cumbrous we shall borrow a summation convention from the tensor calculus. If the affixes i, j range from 1 to p we shall write

$$A^{ij}\,a_{ij} = \sum_{i=1}^{p}\sum_{j=1}^{p} A^{ij}\,a_{ij}, \qquad (28.1)$$

the affixes to A being regarded as ordinary superscripts, not as powers. Similarly we shall have

$$A^{ij}\,a_{ik} = \sum_{i=1}^{p} A^{ij}\,a_{ik}. \qquad (28.2)$$

Whenever an affix occurs as a superscript and a subscript, summation is to be understood. Clearly the actual letter used is a dummy and we have, for instance,

$$A^{ij}\,a_{ij} = A^{kj}\,a_{kj} = A^{kl}\,a_{kl}. \qquad (28.3)$$

We shall write the array of values A^{ij} (a square matrix) as (A^{ij}) and its determinant as $|\,A^{ij}\,|$ or simply as $|\,A\,|$.

To every matrix (a_{ij}) with a non-vanishing determinant there corresponds a reciprocal or inverse matrix which we may write (a^{ij}). Since

$$(a_{ij})\,(a^{ij}) = 1,$$

we have, on carrying out the multiplication,

$$\begin{aligned} a_{ij}\,a^{ik} &= 1, \qquad & j &= k \\ &= 0, & j &\neq k, \end{aligned}$$

which we may express as

$$a_{ij}\,a^{ik} = a_{ij}\,a^{kj} = \delta^k_j, \qquad (28.4)$$

where δ^k_j, one form of the Kronecker delta, is zero if $j \neq k$ and unity otherwise. The quantity a^{ij} is the minor of a_{ij} in $|\,A\,|$ divided by $|\,A\,|$ itself.

28.5. It will further simplify our formulae and will give rise to no loss of generality if we suppose our variates to be in standard measure, that is to say, to have zero mean and unit variance. If we require results for the more general case we can easily obtain them from transformations of the type

$$x_i = \sigma_i\,\xi_i + m_i. \qquad (28.5)$$

With this convention the equation of the multivariate normal distribution (cf. **15.12**, vol. I, p. 376) may be written

$$dF = \frac{\sqrt{|\,A\,|}}{(2\pi)^{\frac{1}{2}p}} \exp\left(-\tfrac{1}{2}A^{ij}\,x_i\,x_j\right) dx_1 \ldots dx_p, \qquad (28.6)$$

where the A's are related to the correlation determinant

$$\Delta = |\,\rho_{ij}\,|. \qquad (28.7)$$

In fact (A^{ij}) is reciprocal to (ρ_{ij}), as we saw in **15.12.**

28.6. We shall also frequently refer to the matrix of sample variances and covariances which we shall call the *dispersion matrix* and write as (a_{ij}), where

$$a_{ij} = \frac{1}{n} \sum_{i,j=1}^{n} (x_i - \bar{x}_i)(x_j - \bar{x}_j). \qquad (28.8)$$

This, it is to be remembered, is in standard measure for the population, that is to say the observed variates are taken from the parent means and divided by the parent standard deviations.

Wishart's Distribution

28.7. We now proceed to generalise to p variates the joint distribution of dispersions arrived at in **14.12** (vol. I, p. 339) for the bivariate case; and we shall also show that the distribution is independent of that of means. The result and method of proof are due to Wishart (1928).

First of all let us write the result for the bivariate case in our new notation. For the distribution of means we have

$$dF = \frac{n\,|\,A\,|^{\frac{1}{2}}}{2\pi} \exp\left(-\frac{n}{2} A^{ij}\,\bar{x}_i\,\bar{x}_j\right) d\bar{x}_1\,d\bar{x}_2, \qquad i, j = 1, 2 \qquad (28.9)$$

and for that of dispersions

$$dF = \left(\frac{n}{2}\right)^{n-1} |\,A\,|^{\frac{1}{2}(n-1)} \frac{|\,a\,|^{\frac{1}{2}(n-4)}}{\pi^{\frac{1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \exp\left(-\frac{n}{2} A^{ij}\,a_{ij}\right) da_{11}\,da_{12}\,da_{22}. \qquad (28.10)$$

For instance, we have

$$a_{11} = s_1^2, \qquad a_{12} = r\,s_1\,s_2, \qquad a_{22} = s_2^2$$

$$(A^{ij}) = \begin{pmatrix} \dfrac{1}{1-\rho^2} & \dfrac{-\rho}{1-\rho^2} \\[2ex] \dfrac{-\rho}{1-\rho^2} & \dfrac{1}{1-\rho^2} \end{pmatrix}$$

so that (28.10) is equivalent to

$$dF = \frac{n^{n-1}}{2^{n-3}\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \frac{(1-r^2)^{\frac{n-4}{2}}\,s_1^{n-2}\,s_2^{n-2}}{(1-\rho^2)^{\frac{1}{2}(n-1)}}$$

$$\times \exp\left\{-\frac{n}{2(1-\rho^2)}(s_1^2 - 2\rho r s_1 s_2 + s_2^2)\right\} ds_1\,ds_2\,dr.$$

This, with the substitution

$$\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right) = \frac{\sqrt{\pi}\,\Gamma(n-2)}{2^{n-3}}$$

is the form found in equation (14.44), vol. I, p. 342, when it is remembered that we are working in standard measure.

28.8. Now consider the general case. With a sample of n values of p variates we consider p rectangular spaces of n dimensions each as the domain of variation. If a point in one of these spaces be fixed, the variation in the other spaces is constrained for fixed values of the sample dispersions. The following argument is a generalisation of that given in **14.12** leading to the bivariate result, and the reader may like to refresh his memory by re-reading that section.

Writing $x_{j1} \ldots x_{jn}$ for the n values of the jth variate, we have for the density function of the whole sample, from (28.6),

$$f = \frac{|A|^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}np}} \exp\left\{-\tfrac{1}{2}\sum_{k=1}^{n}(A^{ij}x_{ik}x_{jk})\right\}$$

$$= \frac{|A|^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}np}} \exp\left[-\tfrac{1}{2}\Sigma\{A^{ij}(x_{ik}-\bar{x}_i)(x_{jk}-\bar{x}_j)\}\right] \times \exp\left(-\frac{n}{2}A^{ij}\bar{x}_i\bar{x}_j\right). \quad (28.11)$$

We may thus factorise the density function into two parts,

$$f_1 = \frac{n^{\frac{1}{2}p}|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp\left(-\frac{n}{2}A^{ij}\bar{x}_i\bar{x}_j\right) \quad . \quad . \quad . \quad . \quad (28.12)$$

and

$$f_2 = \frac{|A|^{\frac{1}{2}(n-1)}}{n^{\frac{1}{2}p}(2\pi)^{\frac{1}{2}(n-1)p}} \exp\left(-\frac{n}{2}A^{ij}a_{ij}\right), \quad . \quad . \quad . \quad (28.13)$$

where we have chosen the constant factor of f_1 so that the distribution shall have the total frequency unity.

Consider now the volume element $\prod_{k=1}^{n} dx_{1k}\,dx_{2k} \ldots dx_{pk}$. In any particular n-space the density is constant over hyperspheres centred at the mean. The volume element may then be represented as the product of elements $d\bar{x}_j$ and of independent elements depending on dispersions. In the total space of pn dimensions the volume element may thus be represented as the product of p elements $d\bar{x}_j$ and an independent element depending on dispersions. Thus the volume element also factorises, and we have immediately for the distribution of means

$$dF = \frac{n^{\frac{1}{2}p}|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp\left(-\frac{n}{2}A^{ij}\bar{x}_i\bar{x}_j\right)\prod_{j=1}^{p} d\bar{x}_j, \quad . \quad . \quad . \quad (28.14)$$

showing that the means are distributed in the multivariate normal form independently of dispersions.

If we define a matrix (B) with elements $\frac{1}{2}n$ times those of (A), we may write the distribution of means in the simple form

$$dF = \frac{|B|^{\frac{1}{2}}}{\pi^{\frac{1}{2}p}} \exp(-B^{ij}\bar{x}_i\bar{x}_j)\,\Pi\,d\bar{x}. \quad . \quad . \quad . \quad . \quad (28.15)$$

We note that this checks with the known results for $p = 1$ and $p = 2$. It is also seen almost at once that the variance of $\bar{x}_j$ is σ_j^2/n, as we expect.

28.9. We have now to consider the more complicated expression for the volume element of dispersions. Let us in the first instance transfer our origins to the sample means, remembering that in doing so we have lost one dimension (or degree of freedom) in the variation of our sample-points. Let $P_1 \ldots P_p$ be the sample-points whose co-ordinates are the n values of $x_1 \ldots x_p$, one point P lying in each n-space. We shall consider in turn the variation of P_1, then that of P_2 for fixed P_1, then that of P_3 for fixed P_1 and P_2,

and so on. The total variation will be given by multiplying the various expressions so obtained; and it will be sufficient if we consider the typical case of the variation of P_m for $m-1$ fixed points $P_1 \ldots P_{m-1}$.

For a fixed length OP_m and fixed angles with $OP_1 \ldots OP_{m-1}$, P_m can vary on a hypersphere of $n-m$ dimensions; for, if we fix any particular angle, P_m is constrained to lie on a hypercone which cuts its hypersphere of variation in a hypersphere of one fewer dimensions, and the fixation of the origin at the sample mean imposes a further constraint. Further, if we regard the p spaces as superposed, as we may, the centre of this $(n-m)$-dimensional hypersphere is the foot of the perpendicular from P_m on to the space containing the points, $O, P_1 \ldots P_{m-1}$. Call the length of this perpendicular for the time being r_m.

The volume of a k-dimensional hypersphere of radius r is

$$\frac{\pi^{\frac{1}{2}k}\, r^k}{\Gamma\left(\frac{k+2}{2}\right)}$$

and its surface area, obtained by differentiating with respect to r, is

$$\frac{2\,\pi^{\frac{1}{2}k}\, r^{k-1}}{\Gamma\left(\frac{1}{2}k\right)}. \qquad (28.16)$$

The surface area of the hypersphere of variation of P_m is thus

$$\frac{2\pi^{\frac{1}{2}(n-m)}\, r_m^{n-m-1}}{\Gamma\left(\frac{n-m}{2}\right)}. \qquad (28.17)$$

To find the element of volume due to the variation of P_m and the angles which OP_m makes with $OP_1 \ldots OP_{m-1}$ we have to multiply (28.17) by an element of variation normal to the hypersphere of $n-m$ dimensions. This variation lies in the hyperplane determined by the origin and $P_1 \ldots P_m$ which is, in fact, normal to the hypersphere. To evaluate it, consider the transformation

$$\xi_{mj} = \sum_{k=1}^{m} x_{mk}\, x_{jk}, \qquad j = 1 \ldots m, \qquad (28.18)$$

where, of course, the x's are measured from the sample means in virtue of our choice of origin. We have for the Jacobian—

$$\begin{aligned} J &= \frac{\partial(\xi_{m1} \ldots \xi_{mm})}{\partial(x_{m1} \ldots x_{mm})} \\ &= \begin{vmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{12} & x_{22} & \ldots & x_{2m} \\ \cdot & \cdot & \cdots & \cdot \\ 2x_{1m} & 2x_{2m} & \ldots & 2x_{mm} \end{vmatrix} \\ &= 2v_m, \end{aligned} \qquad (28.19)$$

where v_m is the volume (or " content ") of the hyperparallelopiped having one corner at the origin and edges running to the points $P_1 \ldots P_m$. Furthermore,

$$\begin{aligned} |\,\xi_{mj}\,| &= |\,\Sigma\, x_{mk}\, x_{jk}\,| \\ &= |\,x_{mk}\,|^2 \\ &= v_m^2. \end{aligned} \qquad (28.20)$$

The required element is thus

$$\frac{1}{2v_m} \prod_{k=1}^{m} d\xi_{mk},$$

and the total element of variation of P_m, on multiplication by (28.17), is

$$\frac{\pi^{\frac{1}{2}(n-m)} \, r_m^{n-m-1}}{\Gamma\left(\frac{n-m}{2}\right) v_m} \prod_{k=1}^{m} d\xi_{mk}. \quad . \quad . \quad . \quad . \quad (28.21)$$

Now r_m is the length of the perpendicular from P_m on to the space $OP_1 \ldots P_{m-1}$ and is therefore equal to v_m/v_{m-1}. Hence, for the variation of P_m we have the element

$$\frac{\pi^{\frac{1}{2}(n-m)}}{\Gamma\left(\frac{n-m}{2}\right)} \frac{v_m^{n-m-2}}{v_{m-1}^{n-m-1}} \prod_{k=1}^{m} d\xi_{mk}. \quad . \quad . \quad . \quad . \quad (28.22)$$

We now derive the total element for variation of $P_1 \ldots P_m$ by multiplying expressions of type (28.22) for $m = 1, 2, \ldots p$. The terms in v cancel except v_p and v_0, the latter being unity, and we find

$$\frac{\pi^{\frac{1}{4}p(2n-p-1)}}{\prod_{k=1}^{p} \Gamma\left(\frac{n-k}{2}\right)} v_p^{n-p-2} \prod_{j=1}^{m} \prod_{k=m}^{p} d\xi_{jk}. \quad . \quad . \quad . \quad (28.23)$$

Now from (28.18) we have

$$\xi_{jk} = n\, a_{jk} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.24)$$

and from (28.20)

$$v_p^2 = n^p \mid a \mid . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.25)$$

Making the necessary substitutions in (28.23) and adjoining the frequency element given by (28.13) we find, after a little reduction,

$$dF = \frac{\left(\frac{n}{2}\right)^{\frac{1}{2}p(n-1)} \mid A \mid^{\frac{1}{2}(n-1)} \mid a \mid^{\frac{1}{2}(n-p-2)}}{\pi^{\frac{1}{4}p(p-1)} \prod_{k=1}^{p} \Gamma\left(\frac{n-k}{2}\right)} \exp\left(-\frac{n}{2} A^{ij} a_{ij}\right) \Pi\, da. \quad . \quad (28.26)$$

This is Wishart's generalisation of the distribution of dispersions in a multivariate normal system. The reader who feels that the foregoing proof demands too much of his powers of geometrical insight may refer to alternative derivations by Wishart and Bartlett (1933*c*) or P. L. Hsu (1939*a*). The domain of variation of the a's is 0 to ∞ for a_{ii} and corresponding values for a_{ij}, $i \neq j$, such that correlations do not exceed unity in absolute value.

28.10. It must be remembered that we are regarding a_{ij} as the same as a_{ji} and that the product of differential elements in (28.26) contains $\frac{1}{2}p\,(p+1)$ items, not p^2; for there are p elements of the form da_{ii} and $\frac{1}{2}p\,(p-1)$ of the form da_{ij}, $i \neq j$. The expanded form of $A^{ij}\, a_{ij}$, however, takes place over i, j from 1 to p, so that any particular term such as $A^{34}\, a_{34}$ occurs twice, once as $A^{34}\, a_{34}$ and once as $A^{43}\, a_{43}$; except that when $i = j$ the term occurs once. For instance, with $p = 2$ we have

$$A^{ij}\, a_{ij} = A^{11}\, a_{11} + 2A^{12}\, a_{12} + A^{22}\, a_{22}. \quad . \quad . \quad . \quad (28.27)$$

We can now derive the characteristic function of the Wishart distribution. Ignoring

constant factors and writing a single integral sign for summation over all a_{ij}, we have, from (28.26)—

$$\int_{-\infty}^{\infty} | a |^{\frac{1}{2}(n-p-2)} \exp\left(-\frac{n}{2} A^{ij} a_{ij}\right) \Pi \, da = \frac{K}{| A |^{\frac{1}{2}(n-1)}} \quad . \quad . \quad (28.28)$$

where K is some constant. In this form let us replace A^{ij} by $A^{ij} - \frac{1}{n}\theta^{ij}$ when $i \neq j$ and by $A^{ij} - \frac{2}{n}\theta^{ij}$ when $i = j$. Then the resulting integral is the characteristic function of the a's, θ^{ij} being the parameter it^{ij} corresponding to a_{ij}. We thus have

$$\phi(\theta^{ij}) = \frac{| A |^{\frac{1}{2}(n-1)}}{\begin{vmatrix} A^{11} - \frac{2}{n}\theta^{11} & A^{12} - \frac{1}{n}\theta^{12} & \dots & A^{1p} - \frac{1}{n}\theta^{1p} \\ A^{12} - \frac{1}{n}\theta^{12} & A^{22} - \frac{2}{n}\theta^{22} & \dots & A^{2p} - \frac{1}{n}\theta^{2p} \\ \cdot & \cdot & \dots & \cdot \\ A^{1p} - \frac{1}{n}\theta^{1p} & A^{2p} - \frac{1}{n}\theta^{2p} & \dots & A^{pp} - \frac{2}{n}\theta^{pp} \end{vmatrix}} \quad . \quad . \quad (28.29)$$

the constant being evaluated by the consideration that $\phi(0) = 1$.

Example 28.1

Let us apply these results to an examination of the moments of the distribution of covariance in the bivariate case. We have

$$A^{11} = A^{22} = \frac{1}{1-\rho^2}, \qquad A^{12} = \frac{-\rho}{1-\rho^2}.$$

We then find for the c. f. of a_{11}, a_{12}, a_{22}—

$$\phi \propto \begin{vmatrix} \frac{1}{1-\rho^2} - \frac{2\theta^{11}}{n} & \frac{-\rho}{1-\rho^2} - \frac{\theta^{12}}{n} \\ \frac{-\rho}{1-\rho^2} - \frac{\theta^{12}}{n} & \frac{1}{1-\rho^2} - \frac{2\theta^{22}}{n} \end{vmatrix}^{-\frac{1}{2}(n-1)}$$

We are interested only in the parameter θ^{12} which we will write as θ, putting the others equal to zero. We then find—

$$\phi \propto \left[\frac{1}{(1-\rho^2)^2} - \left\{\frac{-\rho}{1-\rho^2} - \frac{\theta}{n}\right\}^2\right]^{-\frac{1}{2}(n-1)}$$

$$\phi = \left\{1 - \frac{2\rho\theta}{n} - \frac{(1-\rho^2)\theta^2}{n^2}\right\}^{-\frac{1}{2}(n-1)}$$

Taking logarithms and evaluating coefficients of powers of θ, we find for the cumulants

$$\kappa_1 = \frac{n-1}{n}\rho$$

$$\kappa_2 = \frac{n-1}{n^2}(1+\rho^2)$$

$$\kappa_3 = \frac{2(n-1)}{n^3}\rho(3+\rho^2)$$

$$\kappa_4 = \frac{6(n-1)}{n^4}(1+6\rho^2+\rho^4).$$

In standard measure the distribution tends to normality as n tends to infinity. But for finite n we have

$$\beta_1 = \frac{4}{n-1}\,\frac{\rho^2(3+\rho^2)^2}{(1+\rho^2)^3}$$

$$\beta_2 = 3 + \frac{6}{n-1}\,\frac{1+6\rho^2+\rho^4}{(1+\rho^2)^2}.$$

Thus, even when $\rho = 0$ our distribution, though symmetrical, is not normal.

Wishart (1928) has given formulae as far as those of the fourth order for eight or fewer variates.

Hotelling's Distribution

28.11. In the univariate case we can test the significance of a mean by comparing it with the estimated standard deviation, the ratio being distributed in "Student's" form (or some simple transformation of it if we compare the mean with the actual sample variance and not the unbiassed estimator). We proceed to generalise this result.

We require a single quantity which will serve as a measure of departure of all the means $\bar{x}_j$ from the population values which, as usual, we take to be zero. In place of the matrix of dispersions, we shall consider the matrix of sums of squares and products (b_{ij}) where

$$b_{ij} = \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j). \qquad (28.30)$$

As usual we take (b^{ij}) to be the matrix inverse to (b_{ij}). Let us now write

$$T^2 = n(n-1)\, b^{ij}\,\bar{x}_i\bar{x}_j. \qquad (28.31)$$

This is Hotelling's generalisation of the "Student" ratio t.

In the simplest case when $p = 1$ we have

$$b_{11} = ns^2$$

$$b^{11} = \frac{1}{ns^2},$$

and hence

$$T^2 = \frac{n-1}{s^2}\,\bar{x}^2, \qquad (28.32)$$

so that T becomes equal to the ratio t as required.

28.12. We have

$$\frac{T^2}{n-1} = n\, b^{ij}\,\bar{x}_i\,\bar{x}_j. \qquad (28.33)$$

Let us now denote by m_{ij} the sum of squares or products about the origin, so that

$$m_{ij} = b_{ij} + n\bar{x}_i\bar{x}_j. \qquad (28.34)$$

The determinant of m_{ij} may be written

$$\begin{vmatrix} 1 & \bar{x}_1\sqrt{n} & \bar{x}_2\sqrt{n} & \cdots & \bar{x}_p\sqrt{n} \\ 0 & b_{11}+n\bar{x}_1^2 & b_{12}+n\bar{x}_1\bar{x}_2 & \cdots & b_{1p}+n\bar{x}_1\bar{x}_p \\ 0 & b_{12}+n\bar{x}_2\bar{x}_1 & b_{22}+n\bar{x}_2^2 & \cdots & b_{2p}+n\bar{x}_2\bar{x}_p \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & b_{1p}+n\bar{x}_p\bar{x}_1 & b_{2p}+n\bar{x}_p\bar{x}_2 & \cdots & b_{pp}+n\bar{x}_p^2 \end{vmatrix}$$

On subtracting $\bar{x}_1\sqrt{n}$ times the first row from the second, and so on, we find—

$$| m_{ij} | = \begin{vmatrix} 1 & \bar{x}_1\sqrt{n} & \cdots & \bar{x}_p\sqrt{n} \\ -\bar{x}_1\sqrt{n} & b_{11} & \cdots & b_{1p} \\ \cdot & \cdot & \cdots & \cdot \\ -\bar{x}_p\sqrt{n} & b_{1p} & \cdots & b_{pp} \end{vmatrix}$$

and on expanding according to the border row and column,

$$| m_{ij} | = | b_{ij} | + nb^{ij}\,\bar{x}_i\bar{x}_j\,| b_{ij} |. \quad \quad (28.35)$$

It follows that

$$| b_{ij} | \frac{T^2}{n-1} = | m_{ij} | - | b_{ij} |$$

or

$$\frac{1}{1 + \dfrac{T^2}{n-1}} = \frac{| b_{ij} |}{| m_{ij} |}. \quad \quad (28.36)$$

This is a fundamental equation in the sampling theory of T and we proceed to interpret it geometrically.

28.13. In the case $p = 1$ we have a single sample space of n dimensions. The numerator and denominator of (28.36) then reduce to b_{11} and m_{11}—that is to say, the squares of distances from the sample-point P_1 to its projection on the unit vector whose direction cosines are all equal, and from P_1 to the origin, respectively. The ratio of (28.36) has zero dimensions and is in fact the square of the sine of the angle between OP_1 and the unit vector. This is the geometrical approach which gave us "Student's" distribution in Example 10.6 (vol. I, p. 239).

In the general case let us regard the p n-spaces as superposed in one n-space. The points $P_1 \ldots P_p$ will lie in a space of $p - 1$ dimensions, a hyperplane in the n-space. Now we may rotate the axis without altering the functions $| m_{ij} |$ or $| b_{ij} |$ which are easily seen to be invariant under orthogonal variate-transformations. If we perform such a rotation so as to bring the $(p - 1)$-space of sample-points into correspondence with $p - 1$ co-ordinate dimensions, we see from (28.20) that $| m_{ij} |$ is the square of the content of a hyperparallelopiped with one corner at the origin and sides parallel to $OP_1 \ldots OP_p$.

Now consider a hyperplane perpendicular to the unit vector meeting it, say, in O', and let $P_1' \ldots P_p'$ be the projections of the points P on to this hyperplane. Then b_{ij} is the covariance of the co-ordinates P_i' and P_j' referred to O', and hence $| b_{ij} |$ is the square of the content of the hyperparallelopiped in the hyperplane. Furthermore, the content of this figure bears to that given by $| m_{ij} |$ a ratio equal to the cosine of the angle between the unit vector and the hyperplane. Representing this angle by θ, we have

$$\frac{1}{1 + \dfrac{T^2}{n-1}} = \cos^2\theta. \quad \quad (28.37)$$

28.14. Now if the sample-points P are distributed in the n-space with random orientation, the hyperplane which they determine will be distributed randomly in regard to the angle which it makes with a fixed vector, and in particular with the unit vector. The sampling distribution of θ is then that of an angle between a fixed vector and a random plane. But this, from a slightly different viewpoint, is precisely the problem of distribution which we solved in connection with the multiple correlation coefficient R, for we saw (**15.18**,

vol. I, p. 381) that R is the sine of the angle between a residual vector represented by a variate $x_{1.2\ldots p}$ and the space containing other variates $x_2 \ldots x_p$; and in the case when the former is independent of the latter we can regard it as fixed. Thus, from (28.37) we may write—

$$\frac{1}{1+\dfrac{T^2}{n-1}} = 1 - R^2. \qquad (28.38)$$

The distribution of R^2 in the case when the variate concerned is independent of the others is

$$dF = \frac{1}{B\left(\dfrac{n-p}{2}, \dfrac{p-1}{2}\right)} (1 - R^2)^{\frac{1}{2}(n-p-2)} \; (R^2)^{\frac{1}{2}(p-3)} \, dR^2, \qquad (28.39)$$

where we must remember that p is the *total* number of variates and the variates are measured from their means in forming the regression equation. Before substituting (28.38) in this expression we must increase p by unity, since in effect we are considering $p + 1$ variates —the unit vector determining an additional one; and we must also increase n by unity because our variation is not restricted to that about the mean, as for multiple correlation. With these alterations in (28.39), we have, on substituting for R from (28.38) and a little reduction,

$$dF = \frac{1}{B\left(\dfrac{n-p}{2}, \dfrac{p}{2}\right)} \frac{\{T^2/(n-1)\}^{\frac{1}{2}(p-2)}}{\left(1+\dfrac{T^2}{n-1}\right)^{\frac{1}{2}n}} \, d\left(\frac{T^2}{n-1}\right). \qquad (28.40)$$

This is the distribution of Hotelling's generalisation of "Student's" ratio.

28.15. At the end of the chapter we shall see that this is a particular case of a more general distribution (**28.31**). A third and instructive derivation, due to Wilks, is as follows :—

From the manner of derivation of Wishart's distribution it will be clear that if we substitute the moments about the origin a'_{ij} for those about the mean a_{ij}, the distribution is the same, except that there is an extra degree of freedom. The distribution is then

$$dF = \frac{\left(\dfrac{n^p \mid A \mid}{2^p}\right)^{\frac{1}{2}n} \mid a' \mid^{\frac{1}{2}(n-p-1)}}{\pi^{\frac{1}{4}p(p-1)} \; \Pi \; \Gamma\left(\dfrac{n+1-k}{2}\right)} \exp\left(-\frac{n}{2} A^{ij} a'_{ij}\right) \Pi \, da'.$$

Putting $B^{ij} = \dfrac{n}{2} A^{ij}$, we find, on integration,

$$\int \mid a' \mid^{\frac{1}{2}(n-p-1)} \exp\left(- B^{ij} a'_{ij}\right) \Pi \, da' = \frac{\pi^{\frac{1}{4}p(p-1)} \; \Pi \; \Gamma\left(\dfrac{n+1-k}{2}\right)}{\mid B \mid^{\frac{1}{2}n}}. \qquad (28.41)$$

Now replace n by $n + 2r$ in this expression and divide by the term on the right in (28.41). The result is to give us the rth moment of $\mid a' \mid$ as

$$\mu_r\left(\mid a' \mid\right) = \frac{1}{\mid B \mid^r} \prod_{k=1}^{p} \frac{\Gamma\left(\dfrac{n+1-k}{2} + r\right)}{\Gamma\left(\dfrac{n+1-k}{2}\right)}. \qquad (28.42)$$

We may also write the distribution of a'_{ij} in the form given by our original derivation of Wishart's distribution :—

$$dF = \frac{|B|^{\frac{1}{2}(n-1)} \, |a|^{\frac{1}{2}(n-p-2)}}{\pi^{\frac{1}{4}p(p-1)} \, \Pi \, \Gamma\left(\frac{n-k}{2}\right)} \exp\left(-B^{ij} a_{ij}\right) \Pi \, da \times \frac{|B|^{\frac{1}{2}}}{\pi^{\frac{1}{2}p}} \exp\left(-B^{ij} \bar{x}_i \bar{x}_j\right) \Pi \, d\bar{x}.$$

Multiply this by $|a'|^r$, integrate, and use (28.42), transferring constant terms to the right as in (28.41); then replace n by $n + 2s$ and divide by the constant terms as they were before substitution. We find—

$$\mu'_{r,s}\left(|a'|, |a|\right) = \frac{1}{|B|^{r+s}} \prod_{k=1}^{p} \frac{\Gamma\left(\frac{n+1-k}{2} + r + s\right) \Gamma\left(\frac{n-k}{2} + s\right)}{\Gamma\left(\frac{n+1-k}{2} + s\right) \Gamma\left(\frac{n-k}{2}\right)}. \quad (28.43)$$

Now put $r = -s$ and note that

$$\frac{|a|}{|a'|} = \frac{|b|}{|m|}.$$

We find

$$\begin{aligned} \mu'_s\left(\frac{|b|}{|m|}\right) &= \frac{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-p}{2} + s\right)}{\Gamma\left(\frac{n}{2} + s\right) \Gamma\left(\frac{n-p}{2}\right)} \\ &= \frac{B\left(\frac{n-p}{2} + s, \frac{p}{2}\right)}{B\left(\frac{n-p}{2}, \frac{p}{2}\right)}. \quad (28.44) \end{aligned}$$

Now the function on the right is the sth moment of

$$dF = \frac{1}{B\left(\frac{n-p}{2}, \frac{p}{2}\right)} x^{\frac{1}{2}(n-p-2)} (1-x)^{\frac{1}{2}(p-2)} \, dx \quad . \quad . \quad . \quad (28.45)$$

which is uniquely determined by its moments. This, then, is the distribution of the ratio $\frac{|b|}{|m|}$, and on substitution in terms of T from (28.36) brings us back to the distribution of (28.40). Incidentally this method gives us one more derivation of the distribution of multiple correlations and correlation ratios when the respective variates are independent.

Significance of a Set of Means

28.16. Suppose that we have a set of k samples with numbers $n_1 \ldots n_k$, each from a p-variate population. Let us also suppose that the populations have the same dispersion matrix but different means, that of the jth variate in the lth sample being $\mu_{j(l)}$. We proceed to derive a criterion for testing the means simultaneously. Our result is a generalisation of the testing of k means in normal samples, and we shall obtain it by applying the same method, namely by using the likelihood criterion

$$\lambda = \frac{p_0\,(\omega \text{ max.})}{p_1\,(\Omega \text{ max.})}$$

as given in equation (26.64). Here ω is the domain for which all the means of the jth

variate have a common value μ_j and Ω that for which they have the more general values $\mu_{j\,(l)}$.

Let $b_{ij\,(l)}$ be the function b_{ij} for the lth sample ($l = 1, 2, \ldots k$) and $\bar{x}_{i\,(l)}$ the mean of the ith variate in that sample. Put

$$\bar{b}_{ij} = \sum_{l=1}^{k} b_{ij\,(l)} \qquad . \quad . \quad . \quad . \quad . \quad (28.46)$$

where, of course,

$$b_{ij\,(l)} = \sum_{t=1}^{n_l} (x_{it\,(l)} - \bar{x}_{i\,(l)})\,(x_{jt\,(l)} - \bar{x}_{j\,(l)}). \qquad . \quad . \quad . \quad . \quad . \quad (28.47)$$

Put, for the functions of the pooled samples,

$$\bar{x}_i = \frac{1}{n} \sum_{t,\,l} x_{it\,(l)} = \frac{1}{n} \sum_{l} n_l\, \bar{x}_{i\,(l)} \qquad . \quad . \quad . \quad . \quad . \quad (28.48)$$

$$b_{ij} = \sum_{t} \sum_{l} (x_{it\,(l)} - \bar{x}_i)\,(x_{jt\,(l)} - \bar{x}_j). \qquad . \quad . \quad . \quad . \quad (28.49)$$

If then

$$m_{ij\,(l)} = \sum_{t} (x_{it\,(l)} - \mu_{i\,(l)})\,(x_{jt\,(l)} - \mu_{j\,(l)}) \qquad . \quad . \quad . \quad (28.50)$$

the likelihood of all samples together is

$$c \mid A \mid^{\frac{1}{2}n} \exp \left\{ - \tfrac{1}{2} \sum_{l} (n_l\, A^{ij}\, m_{ij\,(l)}) \right\}, \qquad . \quad . \quad . \quad . \quad (28.51)$$

where c is a constant.

Taking logarithms and differentiating, we have for the maximum value equations typified by

$$\sum_{l} \sum_{t} n_l\, A^{ij} \left\{ (x_{it\,(l)} - \mu_{i\,(l)}) + (x_{jt\,(l)} - \mu_{j\,(l)}) \right\} = 0,$$

which reduce to

$$\bar{x}_{i\,(l)} = \hat{\mu}_{i\,(l)}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (28.52)$$

The maximum likelihood values of the m's are then given by

$$\hat{m}_{ij\,(l)} = b_{ij\,(l)}.$$

Furthermore, the values of $\hat{A}^{ij}$ are then given by the inverse of the matrix $\left(\frac{1}{n}\,\bar{b}_{ij}\right)$, and the exponent of (28.51) becomes

$$- \tfrac{1}{2}n \sum (\hat{A}^{ij}\, b_{ij\,(l)}) = - \tfrac{1}{2}nk. \qquad . \quad . \quad . \quad . \quad (28.53)$$

We then find

$$p_1\,(\Omega \text{ max.}) = \frac{c\, e^{-\frac{1}{2}nk}}{\left| \frac{1}{n}\, \bar{b}_{ij} \right|^{\frac{1}{2}n}}. \qquad . \quad . \quad . \quad . \quad (28.54)$$

In a similar way it will be found that

$$p_0\,(\omega \text{ max.}) = \frac{c\, e^{-\frac{1}{2}nk}}{\left| \frac{1}{n}\, b_{ij} \right|^{\frac{1}{2}n}}. \qquad . \quad . \quad . \quad . \quad (28.55)$$

Hence

$$\lambda = \frac{\left|\frac{1}{n}\bar{b}_{ij}\right|^{\frac{1}{2}n}}{\left|\frac{1}{n}b_{ij}\right|^{\frac{1}{2}n}}$$

and we may write

$$L = \lambda^{\frac{2}{n}} = \frac{\left|\frac{1}{n}\bar{b}_{ij}\right|}{\left|\frac{1}{n}b_{ij}\right|} = \frac{|\bar{b}_{ij}|}{|b_{ij}|} \quad . \quad . \quad . \quad . \quad . \quad (28.56)$$

and take L as our criterion.

28.17. The distribution of L for general k is not easily expressible, but we may determine its moments by the method employed in **28.15.** The functions $\frac{1}{n}b_{ij}$ are distributed in Wishart's form and their moments accordingly given by equations of the type (28.42) with n replaced by $n-1$, namely,

$$\mu_r'(|b_{ij}|) = \frac{n^{pr}}{|B|^r}\prod_{m=1}^{p}\frac{\Gamma\left(\frac{n-m}{2}+r\right)}{\Gamma\left(\frac{n-m}{2}\right)} \quad . \quad . \quad . \quad . \quad (28.57)$$

Now each $b_{ij\,(l)}$ is distributed in Wishart's form, and therefore their sum is so distributed (cf. Exercise 28.3). In the manner of **28.15**—we omit the details—it is found that

$$\mu_r'\left(\frac{|\bar{b}_{ij}|}{|b_{ij}|}\right) = \prod_{m=1}^{p}\frac{\Gamma\left(\frac{n-m}{2}\right)\Gamma\left(\frac{n-m+1-k}{2}+r\right)}{\Gamma\left(\frac{n-m}{2}+r\right)\Gamma\left(\frac{n-m+1-k}{2}\right)} \quad . \quad . \quad (28.58)$$

where we now use m as an index of summation, reserving k for the number of samples. This gives us the moments of L.

In the case $k = 2$ we have

$$\mu_r' = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-p-1}{2}+r\right)}{\Gamma\left(\frac{n-1}{2}+r\right)\Gamma\left(\frac{n-p-1}{2}\right)} \quad . \quad . \quad . \quad . \quad (28.59)$$

and hence the distribution of L is in the form

$$dF = \frac{1}{B\left(\frac{n-p-1}{2}, \frac{p}{2}\right)} L^{\frac{1}{2}(n-p-3)}\,(1-L)^{\frac{1}{2}(p-2)}\,dL. \quad . \quad . \quad (28.60)$$

In the case $k = 3$ we find

$$\mu_r' = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-p-1}{2}+r\right)\Gamma\left(\frac{n-p-2}{2}+r\right)}{\Gamma\left(\frac{n-1}{2}+r\right)\Gamma\left(\frac{n-2}{2}+r\right)\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{n-p-2}{2}\right)}$$

which, in virtue of the relation

$$\Gamma(x+\tfrac{1}{2})\,\Gamma(x+1) = \frac{\sqrt{\pi}\,\Gamma(2x+1)}{2^{2x}}$$

becomes

$$\mu_r' = \frac{\Gamma(n-2)\,\Gamma(n-p-2+2r)}{\Gamma(n-2+2r)\,\Gamma(n-p-2)} \quad . \quad . \quad . \quad . \quad (28.61)$$

These are the moments of the distribution

$$dF = \frac{1}{2B(n-p-2,\,p)}(\sqrt{L})^{n-p-4}(1-\sqrt{L})^{p-1}\,dL, \quad . \quad . \quad (28.62)$$

a rather unusual form. The results are due to Wilks.

28.18. The line of generalisation of univariate analysis will now probably be clear. Corresponding to most of our results for a single variate there will be a generalised result for p variates ; and, in fact, if we like to regard the p-variate as a vector we can often draw direct analogies between results for vectors and those for the (univariate) scalar. It is of special interest to observe that the role played by the variance in univariate theory is taken over by the determinant of the dispersion matrix in multivariate theory.

Up to this point we have generalised the distribution of variance (the χ^2-distribution) into Wishart's form, and the t-distribution into Hotelling's form.

Other results which suggest themselves for generalisation are regression and variance analysis. But in a sense our treatment of regressions is already general, for we have discussed the regression of one variate on $p-1$ others. Below we shall go further and examine the relations between p dependent and q independent variates. In vector language, we consider the regression of a p-way vector y on a q-way vector x. We have also considered the analysis of variance for the bivariate and trivariate case in Chapter 24 under the title of analysis of covariance, and since the interest lies mainly in the direction of regressions we shall not take the subject further here, though it is capable of development and even, perhaps, of application if data become available in sufficient abundance. In the remainder of the chapter we shall, in the first instance, deal with an offshoot of regression theory which has some interesting taxonomic applications, namely discriminatory analysis ; and we shall then proceed to the general problem of the relationship between two sets of variates.

Discriminatory Analysis

28.19. Suppose we have p observations for each of $2n$ sample members, and that each member can have emanated from one of two populations, n to each population. We require to find some measurement depending on the p observations which will enable us to assign subsequently drawn members correctly to their parent populations with the greatest assurance of success. For this purpose we shall find p quantities $\lambda^1 \ldots \lambda^p$ and a *discriminant function* X related linearly to the variates by

$$X = \lambda^j x_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.63)$$

The criterion on which we shall rely is that the λ's must be chosen to maximise the ratio of the difference between sample means to the standard deviation within the two classes.

Any linear function of type (28.63) has variance S, given by

$$S = \lambda^i \lambda^j a_{ij}\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.64)$$

where, as usual, a_{ij} is the covariance of x_i and x_j which we assume to be the same for both populations. Further, if the difference of the two means of x_j is d_j, the difference of the function X for the two samples is

$$D = \lambda^i d_i. \qquad (28.65)$$

We have then to maximise for variation in λ the function

$$\frac{D^2}{S} = \frac{(\lambda^i d_i)^2}{\lambda^i \lambda^j a_{ij}}. \qquad (28.66)$$

This gives for each λ

$$\frac{1}{2}\frac{\partial S}{\partial \lambda} = \frac{S}{D}\frac{\partial D}{\partial \lambda},$$

leading to equations typified by

$$\lambda^j a_{ij} = \frac{S}{D} d_i. \qquad (28.67)$$

Multiplying by a^{ik} and summing over i, we have

$$\lambda^j a_{ij} a^{ik} = \frac{S}{D} d_i a^{ik}$$

$$= \lambda^j \delta^k_j = \lambda^k ;$$

or, replacing k by j,

$$\lambda^j = \frac{S}{D} d_i a^{ij}. \qquad (28.68)$$

This determines the λ's, except for the constant $\frac{S}{D}$ which can be chosen at will so far as the discriminant function is concerned. If c is some constant, we have

$$\lambda^j = c\, d_i a^{ij}. \qquad (28.69)$$

The result also holds if there are n_1 members in the first sample and n_2 in the second. Equation (28.65) remains true, and the rest of the analysis is the same as for equal class-numbers.

Example 28.2 (from R. A. Fisher, 1936*a*).

Measurements were made on fifty specimens of flowers from each of two species of iris, *setosa* and *versicolor*, found growing in the same colony. Four measurements were taken, viz. sepal length, sepal width, petal length, and petal width. We denote them by x_1, x_2, x_3 and x_4 respectively.

The means of the specimens were (in centimetres) :—

Variate.	*Versicolor.*	*Setosa.*	Difference $(V-S)$.
x_1	5·936	5·006	0·930
x_2	2·770	3·428	− 0·658
x_3	4·260	1·462	2·798
x_4	1·326	0·246	1·080

The sums of squares and products about the means were (in cm.^{2}) :—

	x_1	x_2	x_3	x_4
x_1	19·1434	9·0356	9·7634	3·2394
x_2	9·0356	11·8658	4·6232	2·4746
x_3	9·7634	4·6232	12·2978	3·8794
x_4	3·2394	2·4746	3·8794	2·4604

The inverse matrix is, in cm.$^{-2}$:—

	x_1	x_2	x_3	x_4
x_1	0·118,7161	− 0·066,8666	− 0·081,6158	0·039,6350
x_2	− 0·066,8666	0·145,2736	0·033,4101	− 0·110,7529
x_3	− 0·081,6158	0·033,4101	0·219,3614	− 0·272,0206
x_4	0·039,6350	− 0·110,7529	− 0·272,0206	0·894,5506

We need not bother to divide these quantities by n because there is an arbitrary constant in our discriminant function which absorbs it. The matrices are diagonally symmetric, and it is not always necessary to write out the values below the diagonal as we have done here.

From (28.69), with $c = 1$, we then find—

$$\lambda^1 = -0{\cdot}031{,}1511 \qquad \lambda^2 = -0{\cdot}183{,}9075$$
$$\lambda^3 = 0{\cdot}222{,}1044 \qquad \lambda^4 = 0{\cdot}314{,}7370.$$

If we choose the coefficient of x_1 to be unity the discriminant function is then

$$X = x_1 + 5{\cdot}9037x_2 - 7{\cdot}1299x_3 - 10{\cdot}1036x_4. \qquad (28.70)$$

The mean of X for *versicolor*, obtained by substituting the means of the x's for that species, is found to be $-$ 21·4815, and that for *setosa* is 12·3345. The difference is thus 33·816 cm. Let us compare this with its standard error to see whether it is significant of real differences in the values of X for the two species.

From the matrix of sums of squares and products we find

$$N \text{ var } X = \lambda^i \lambda^j a_{ij} = 1085{\cdot}5522,$$

where the λ's are, of course, the coefficients in (28.70). N here is the number of degrees of freedom of the estimate of the variance. There are 100 members altogether, with 99 degrees of freedom, but we have eliminated four corresponding to the means of the four variates. We therefore take N to be $99 - 4 = 95$, and find

$$\text{var } X = 11{\cdot}4269.$$

This is the variance of a single value. That of the difference of the two means of 50 values is obtained by division by 25 and is thus 0·4571, the corresponding standard error being 0·676.

The observed difference of means, viz. 33·816, is about 50 times this amount, and there is thus a real difference in the values of X for the two species. In other words the discriminant function is a good one. It is best among the linear functions of the x's because

we have chosen it so that the difference of two values, divided by their estimated standard error, shall be the greatest possible. To use the function we should, given a flower of doubtful species, calculate X for it and assign it to one species or the other according as X were nearer to the mean value of X for one species or the other. If, of course, the observed value differed from the mean values by more than twice the standard error of each, we should begin to doubt whether it belonged to either.

The analysis may be put in rather a different way. Suppose we analyse the variation of X between and within species. The sum of squares between species in the 50×2 classification is

$$50 \{ (\bar{X}_1 - \bar{X})^2 + (\bar{X}_2 - \bar{X})^2 \},$$

where $\bar{X}_1, \bar{X}_2$ are the respective means and $\bar{X}$ the mean of the whole. This reduces to $25D^2$. The sum of squares within classes is 1085·55 with 95 d.f., as found above, and we have—

Sum of Squares.		d.f.
Between species	28,588·05	4
Within species	1,085·55	95
TOTALS	29,673·60	99

Our method of selecting the discriminant function has been such as to minimise the sum of squares within species and, for constant total, to maximise the sum between species, and hence to minimise the ratio of the latter to the former. For the moment we cannot assume that this ratio may be tested in the z-distribution in the usual way, though we shall see presently that this is so.

28.20. The relationship of discriminatory analysis for two classes and the theory of regression may be brought out by introducing a formal variate y for the classes. If there are n_1 members in one class and n_2 in the other we shall assign the values

$$\frac{n_2}{n_1 + n_2}, \qquad \frac{-n_1}{n_1 + n_2}$$

to the y-variate for the two classes respectively. The mean of y for the whole sample is then zero and the sum of squares is

$$\frac{n_1 n_2}{n_1 + n_2} = \zeta, \text{ say.} \qquad (28.71)$$

Considering now

$$Y = \lambda^j x_j \qquad (28.72)$$

as a regression equation, we find for the coefficients λ

$$\Sigma (Yx_j) - \lambda^i \Sigma (x_i x_j) = 0,$$

or

$$\Sigma (Yx_j) - \lambda^i a_{ij} = 0. \qquad (28.73)$$

Now

$$\Sigma (Yx_j) = \frac{n_2}{n_1 + n_2} \Sigma_1 (x_j) - \frac{n_1}{n_1 + n_2} \Sigma_2 (x_j),$$

where the suffixes of the Σ's relate to the first and second classes,

$$= \frac{n_2 n_1}{n_1 + n_2} \{ (\bar{x}_j)_1 - (\bar{x}_j)_2 \}$$

$$= \zeta d_j.$$

Thus
$$\zeta d_j = \lambda^i a_{ij}, \quad \ldots \ldots \quad (28.74)$$

which is another way of writing (28.69) with a particular value for the constant c.

28.21. Pursuing the analogy with regression analysis further, we see that since

$$\Sigma (Y^2) = \zeta$$

and
$$\Sigma (Yx_j) = \zeta d_j$$

we may analyse the sums of squares as—

Sums of squares.	d.f.
$\zeta \lambda^i d_i$	p
$\zeta (1 - \lambda^i d_i)$	$n_1 + n_2 - p - 1$
ζ	$n_1 + n_2 - 1$ . . . . (28.75)

as for a regression line. If R is the multiple regression of Y on the x-variates,

$$R^2 = \lambda^i d_i. \quad \ldots \ldots \quad (28.76)$$

In ordinary regression analysis we may test the ratio $R^2/(1 - R^2)$, multiplied by suitable constants, in the z-distribution ; but this depends on the assumption that the dependent variate y is normal for any fixed x's. Here we have the case when the dependent variate is fixed but the x's are normal. The test still holds in such a case, the reason being the kind of duality we noted in **28.14** in arriving at Hotelling's distribution. The distribution of angles between a fixed plane and a random vector is the same as that between a fixed vector and a random plane. Consequently the table of (28.75) can be regarded as an analysis of variance and the z-test applied.

28.22. We may extend the discriminant function to the case when the property to be discriminated is not, as above, a matter of allocation to one of two classes, but to several which may in particular be determined by certain values of a continuous variate. If we have various measurements of p x-variates corresponding to values of a y-variate, we may form the regression of y on the x's and use the resulting function as a discriminator. As in the case of dichotomy, the regression will maximise the difference between classes as compared with intra-class variation ; and its significance may be tested in much the same way.

Example 28.3 (from M. M. Barnard, 1935).

An investigation was undertaken into the changes taking place over time of the characteristics of certain Egyptian skulls. There were four sets of skulls, known to be from Late Predynastic, Sixth to Twelfth, Twelfth to Thirteenth and Ptolemaic dynasties respect-

ively, and the relative time-intervals were taken to be in the proportions 2 : 1 : 2, so that the values of t for the four periods may be taken to be respectively -5, -1, $+1$, $+5$.

For the skulls four measurements were selected:

x_1, basi-alveolar length;
x_2, nasal height;
x_3, maximum breadth;
x_4, basi-bregmatic height.

It is required to find a function

$$X = \lambda^1 x_1 + \lambda^2 x_2 + \lambda^3 x_3 + \lambda^4 x_4$$

which will best discriminate between skulls belonging to different periods.

The means of the series were as follows, the sample numbers also being shown:—

Variate.	Series I ($n_1 = 91$).	Series II ($n_2 = 162$).	Series III ($n_3 = 70$).	Series IV ($n_4 = 75$).
x_1	133·582,418	134·265,432	134·371,429	135·306,667
x_2	98·307,692	96·462,963	95·857,143	95·040,000
x_3	50·835,165	51·148,148	50·100,000	52·093,333
x_4	133·000,000	134·882,716	133·642,857	131·466,667

The sums of squares and products about the means are—

	x_1	x_2	x_3	x_4
x_1	9661·997,470	445·573,301	1130·623,900	2148·584,219
x_2	...	9073·115,027	1239·221,990	2255·812,722
x_3	...	...	3938·320,351	1271·054,662
x_4	...	...	...	8741·508,829

The mean value of t, $\bar{t}$, for the 398 observations is $-$ 0·432,161, and the values of $t - \bar{t}$ for the four series are accordingly

$$-4{\cdot}567{,}839;\quad -0{\cdot}567{,}839;\quad 1{\cdot}432{,}161;\quad 5{\cdot}432{,}161.$$

The sums $\Sigma x_j (t - \bar{t})$ are respectively

x_1	718·762,86
x_2	$-$ 1407·260,75
x_3	410·101,94
x_4	$-$ 733·668,32

and finally, $\Sigma (t - \bar{t})^2 = 4307{\cdot}668{,}32$.

We could obtain the coefficients λ from the reciprocal of the matrix above on the lines of the previous example. It is also instructive to observe, from the analogy with regressions, that instead of that matrix we may use the matrix (depending on one extra degree of freedom, 395 in all) obtained by adding to the sums of squares the regressions on time.

For instance, instead of 9661·997,470 we have 9661·997,470 + $(718{\cdot}762{,}86)^2/4307{\cdot}668{,}32$. The resulting matrix is

	x_1	x_2	x_3	x_4
x_1	9781·927,828	210·762,489	1199·052,135	2026·206,952
x_2	...	9532·849,476	1105·246,827	2405·414,318
x_3	...	...	3977·363,203	1201·230,304
x_4	...	...	...	8866·382,928

The reciprocal of this is (units = 10^{-6})—

	x_1	x_2	x_3	x_4
x_1	110·368,975	6·938,481	− 28·145,236	− 23·361,935
x_2	...	115·693,529	− 24·948,984	− 30·767,069
x_3	...	...	273·988,409	− 23·666,591
x_4	...	...	...	129·990,069

The resulting values of λ are

$$\lambda^1 = 0{\cdot}075{,}156{,}739, \qquad \lambda^2 = -\,0{\cdot}145{,}490{,}050,$$
$$\lambda^3 = 0{\cdot}144{,}600{,}884, \qquad \lambda^4 = -\,0{\cdot}078{,}538{,}419$$

and these, or constant multiples of them, give us the constants in the discriminant function which will best enable us to assign a skull to the correct period by measurements of the four specified variates.

In this analysis we have 398 members, but of the 397 d.f. we have discarded two with the general mean. The d.f. of the sum $4307{\cdot}6683 = \Sigma\,(t - \bar{t})^2$ are 395, of which four are attributable to regressions on the other variates. For the contribution of these four we have

$$\lambda^1 \times 718{\cdot}762{,}86 + \text{etc.} = 375{\cdot}6657.$$

The analysis of variance is thus—

Sum of Squares.		d.f.	Quotient.
Regression	375·6657	4	...
Remainder	3932·0026	391	10·0563
Totals . . .	4307·6683	395	

The analogy of the discriminant function with regressions noted above may be used to provide standard errors of the coefficients λ. In our present case the variance of λ^1 is obtained by multiplying the remainder quotient, viz. 10·0563, by the term corresponding

to x_1^2 in the reciprocal matrix of sums of squares of the x's, namely $110{\cdot}368{,}975 \times 10^{-6}$. This gives a standard error of 0·0333. We obtain finally

$$\begin{aligned} \lambda^1 &= 0{\cdot}0752 \pm 0{\cdot}0333 \\ \lambda^2 &= -0{\cdot}1455 \pm 0{\cdot}0341 \\ \lambda^3 &= 0{\cdot}1446 \pm 0{\cdot}0525 \\ \lambda^4 &= -0{\cdot}0785 \pm 0{\cdot}0362. \end{aligned}$$

All coefficients exceed twice their standard error, and hence all the variates are useful in discriminating between skulls of different periods.

I am indebted to Dr. M. S. Bartlett for the calculations of this example. His results differ from those reached by Miss Barnard in her original investigation since she took an unweighted regression of the variates with time, whereas he has weighted the values according to sample numbers. He also notes that the significance of the results has been tested above on the basis of variability within classes, but that a fuller analysis of the means, bringing back the two degrees of freedom discarded, reveals further differences between the series. Thus, though the discriminant function will efficiently sort the series examined in relation to their periods, we must be cautious about associating the observed differences with the time-changes.

Canonical Correlations

28.23. We now turn to consider the general theory of the relations between two sets of variates $x_1 \ldots x_p$ and $x_{p+1} \ldots x_{p+q}$, where we suppose that $p \leqslant q$. Following Hotelling (1936*b*), we shall show that in general there can be found linear transformations to variates $\xi_1 \ldots \xi_p$, $\xi_{p+1} \ldots \xi_{p+q}$ such that

(*a*) all the ξ's have unit variance and zero mean ;

(*b*) any ξ in the p-group is independent of the other ξ's in that group ;

(*c*) any ξ in the q-group is independent of the other ξ's in that group ;

(*d*) the correlation between any ξ in the p-group and any ξ in the q-group is zero except for p correlations $\rho_1 \ldots \rho_p$, which may be taken to be the correlations between ξ_1 and ξ_{p+1}, ξ_2 and ξ_{p+2}, $\ldots$ ξ_p and ξ_{2p}.

The variates ξ are then said to be *canonical variates* and the ρ's *canonical correlations*.

This part of our work is, fundamentally, the reduction of two quadratic forms and an associated bilinear form to canonical types and does not depend on the distribution laws of the variates. Furthermore, the reduction can be carried out either on the population or on the sample. In the latter case it will yield sample canonical correlations which may be written $r_1 \ldots r_p$ and regarded as sample-values of the parent ρ's.

We will suppose that our variates x have zero means and dispersions denoted by σ_{ij}, where, for the time being, we use σ to denote a variance or covariance instead of the more usual σ^2. Those dispersions in the p-group we denote by Greek affixes : $\sigma_{\alpha\beta}$, and those in the q-group by Roman affixes : σ_{ij}. For a covariance of a p-variate with a q-variate we write one Greek and one Roman affix : $\sigma_{\alpha i}$.

Consider now a particular pair of variates given by

$$\left.\begin{aligned} \xi &= c^\alpha x_\alpha, \qquad \alpha = 1, \ldots p \\ \eta &= d^a x_a, \qquad a = 1, \ldots q \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (28.77)$$

If their variances are unity we have

$$\left.\begin{aligned} c^\alpha c^\beta \sigma_{\alpha\beta} &= 1 \\ d^a d^b \sigma_{ab} &= 1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (28.78)$$

We will also impose the condition that their correlation R is stationary for variations in the coefficients c and d, i.e. that

$$R = c^\alpha d^a \sigma_{\alpha a} = \text{stationary.} \qquad (28.79)$$

Equations (28.78) and (28.79) then require an unconditioned stationary value of

$$c^\alpha d^a \sigma_{\alpha a} - \tfrac{1}{2}\lambda c^\alpha c^\beta \sigma_{\alpha\beta} - \tfrac{1}{2}\mu d^a d^b \sigma_{ab} \qquad (28.80)$$

where λ and μ are undetermined multipliers. This leads to

$$\left.\begin{aligned} c^\alpha \sigma_{\alpha a} - \mu d^b \sigma_{ab} &= 0 \\ d^a \sigma_{\alpha a} - \lambda c^\beta \sigma_{\alpha\beta} &= 0 \end{aligned}\right\}. \qquad (28.81)$$

Multiplying the first equation by d^a and summing and the second by c^α and summing, we have, in virtue of (28.78) and (28.79),

$$R = \lambda = \mu. \qquad (28.82)$$

Equations (28.81) will then be soluble for the $p + q$ unknowns c and d if the determinant of their array vanishes, that is if, writing λ for the constants μ and λ,

$$\begin{vmatrix} \lambda\sigma_{11} & \cdots & -\lambda\sigma_{1p} & \sigma_{1,p+1} & \cdots & \sigma_{1,p+q} \\ \cdot & \cdots & & \cdot & \cdots & \cdot \\ \lambda\sigma_{p1} & \cdots & -\lambda\sigma_{pp} & \sigma_{p,p+1} & \cdots & \sigma_{p,p+q} \\ \sigma_{p+1,1} & \cdots & \sigma_{p+1,p} & -\lambda\sigma_{p+1,p+1} & \cdots & -\lambda\sigma_{p+1,p+q} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \sigma_{p+q,1} & \cdots & \sigma_{p+q,p} & -\lambda\sigma_{p+q,p+1} & \cdots & -\lambda\sigma_{p+q,p+q} \end{vmatrix} = 0 \qquad (28.83)$$

an equation determining λ. Before studying it further we will throw the equation into a somewhat different form.

28.24. We may write (28.83) as

$$\left|\begin{array}{c|c} -\lambda\sigma_{\alpha\beta} & \sigma_{\alpha j} \\ \hline \sigma_{i\beta} & -\lambda\sigma_{ij} \end{array}\right| = 0. \qquad (28.84)$$

Multiplying the first p rows by $-\lambda$ and dividing the last q columns by $-\lambda$ we find the equivalent form

$$(-\lambda)^{q-p} \left|\begin{array}{c|c} \lambda^2\sigma_{\alpha\beta} & \sigma_{\alpha j} \\ \hline \sigma_{i\beta} & \sigma_{ij} \end{array}\right| = 0. \qquad (28.85)$$

Writing, in conformity with our usual notation, (σ^{ij}) for the matrix inverse to (σ_{ij}) and remembering that

$$\sigma^{ij} \sigma_{ik} = \delta^j_k,$$

let us multiply (28.85) on the left by

$$\left|\begin{array}{c|c} \delta^\alpha_\gamma & -\sigma^{ik}\sigma_{\gamma k} \\ \hline 0 & \sigma^{li} \end{array}\right|. \qquad (28.86)$$

The product of determinants is then

$$\left|\begin{array}{c:c} \lambda^2 \sigma_{\beta\gamma} - \sigma_{i\beta}\,\sigma^{ik}\,\sigma_{\gamma k} & \delta^{\alpha}_{\gamma}\,\sigma_{\alpha j} - \sigma^{ik}\,\sigma_{\gamma k}\,\sigma_{ij} \\ \hdashline \sigma^{li}\,\sigma_{i\beta} & \sigma^{li}\,\sigma_{ij} \end{array}\right|$$

$$= \left|\begin{array}{c:c} \lambda^2 \sigma_{\beta\gamma} - \sigma_{i\beta}\,\sigma^{ik}\,\sigma_{\gamma k} & 0 \\ \hdashline \sigma^{li}\,\sigma_{i\beta} & \delta^{l}_{j} \end{array}\right|$$

which gives

$$(-\lambda)^{q-p} \mid \lambda^2 \sigma_{\beta\gamma} - \sigma_{i\beta}\,\sigma^{ik}\,\sigma_{\gamma k} \mid = 0, \qquad . \quad . \quad . \quad (28.87)$$

a determinant with p rows and columns multiplied by a power of λ.

28.25. Returning now to our original problem, we see that if a simple root of (28.83) is substituted in (28.81) the c's and d's are determinate, except of course that they may be replaced by $-c$ and $-d$. For a root of multiplicity m they are determinate except for $m-1$ assignable constants, a result we take without proof from the theory of algebraic forms (reference may be made to Hotelling's paper for details).

From (28.87) we see that the equation in λ has $p+q$ roots. It cannot have fewer, for the coefficient of the highest power of λ in (28.83) is the product of two principal minors which do not vanish unless the variates are linearly dependent, a case which we exclude from the discussion. Of these $p+q$ roots $q-p$ are zero. The remaining $2p$ can be grouped in pairs, each of which is the negative of the other. There are thus roots which we may write $\pm \rho_1, \ldots \pm \rho_p$. We choose as the roots those which are not negative and proceed to prove that they are the canonical correlations as we have defined them. That they are, in fact, correlations follows from (28.82).

Suppose we have a root ρ_γ and determine the corresponding constants c_γ and d_γ and hence a pair of variates ξ_γ and η_γ. Then we have, from (28.81),

$$\left.\begin{aligned} c^{\alpha}_{\gamma}\,\sigma_{\alpha a} &= \rho_\gamma\, d^{b}_{\gamma}\,\sigma_{ab} \\ d^{a}_{\gamma}\,\sigma_{\alpha a} &= \rho_\gamma\, c^{\beta}_{\gamma}\,\sigma_{\alpha\beta} \end{aligned}\right\}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (28.88)$$

Similar equations obtain for a second pair, say ξ_δ and η_δ. Between these four variates there are six correlations, two of which are ρ_γ and ρ_δ. We wish to show that the other four vanish. They are

$$\begin{aligned} E(\xi_\gamma, \xi_\delta) &= c^{\alpha}_{\gamma}\, c^{\beta}_{\delta}\,\sigma_{\alpha\beta} & \qquad E(\eta_\gamma\, \eta_\delta) &= d^{a}_{\gamma}\, d^{b}_{\delta}\,\sigma_{ab} \\ E(\xi_\gamma\, \eta_\delta) &= c^{\alpha}_{\gamma}\, d^{b}_{\delta}\,\sigma_{\alpha b} & \qquad E(\xi_\delta\, \eta_\gamma) &= c^{\beta}_{\delta}\, d^{a}_{\gamma}\,\sigma_{a\beta}. \end{aligned} \qquad . \quad . \quad . \quad (28.89)$$

Multiply the first of (28.88) by d^{a}_{δ} and sum. Using (28.89), we have

$$E(\xi_\gamma\, \eta_\delta) = \rho_\gamma\, E(\eta_\gamma\, \eta_\delta). \qquad . \quad . \quad . \quad . \quad . \quad . \quad (28.90)$$

Similarly from the second of (28.88) multiplied by c^{α}_{δ},

$$E(\xi_\delta\, \eta_\gamma) = \rho_\gamma\, E(\xi_\gamma\, \xi_\delta). \qquad . \quad . \quad . \quad . \quad . \quad . \quad (28.91)$$

Interchanging γ and δ we find from (28.90) and (28.91)

$$\rho_\gamma\, E(\eta_\gamma\, \eta_\delta) = \rho_\delta\, E(\xi_\gamma\, \xi_\delta). \qquad . \quad . \quad . \quad . \quad . \quad (28.92)$$

Equally, again interchanging γ and δ in (28.92) we have

$$\rho_\delta\, E(\eta_\gamma\, \eta_\delta) = \rho_\gamma\, E(\xi_\gamma\, \xi_\delta). \qquad . \quad . \quad . \quad . \quad . \quad (28.93)$$

Thus, unless $\rho_\gamma^2 = \rho_\delta^2$,

$$E(\xi_\gamma \xi_\delta) = E(\eta_\gamma \eta_\delta) = 0. \quad . \quad . \quad . \quad . \quad (28.94)$$

It follows from (28.90) and (28.91) that the other correlations also vanish.

We have only to round off the proof by showing that if ρ is a root of multiplicity m the property still holds. This follows from the consideration that we may then choose our c's and d's to obey certain orthogonal conditions ensuring that

$$E(\xi_\gamma \xi_\delta) + E(\eta_\gamma \eta_\delta) = 0. \quad . \quad . \quad . \quad . \quad (28.95)$$

It will then follow from (28.92) that each expectation vanishes unless $\rho_\gamma = \rho_\delta = 0$; and even in this case, (28.91) and (28.92) show that two expectations vanish, and we may then choose our assignable constants so that the others vanish.

28.26. When the variates are put into canonical form the dispersion matrix reduces to

$$\begin{pmatrix}
1 & 0 & . & . & 0 & \rho_1 & 0 & . & . & 0 & . & . & 0 \\
0 & 1 & . & . & 0 & 0 & \rho_2 & . & . & 0 & . & . & 0 \\
. & . & . & . & . & . & . & . & . & . & . & . & . \\
0 & 0 & . & . & 1 & 0 & 0 & . & . & \rho_p & . & . & 0 \\
\rho_1 & 0 & . & . & 0 & 1 & 0 & . & . & 0 & . & . & 0 \\
0 & \rho_2 & . & . & 0 & 0 & 1 & . & . & 0 & . & . & 0 \\
. & . & . & . & . & . & . & . & . & . & . & . & . \\
0 & 0 & . & . & \rho_p & 0 & 0 & . & . & 1 & . & . & 0 \\
. & . & . & . & . & . & . & . & . & . & . & . & . \\
0 & 0 & . & . & 0 & 0 & 0 & . & . & 0 & . & . & 1
\end{pmatrix} \quad . \quad . \quad (28.96)$$

with a determinant equal to

$$(1 - \rho_1^2)(1 - \rho_2^2) \ . \ . \ . \ (1 - \rho_p^2).$$

Example 28.4 (from Hotelling, 1936*b*, dealing with data of T. L. Kelley).

140 seventh-grade school children were given four tests in (*a*) reading speed, (*b*) reading power, (*c*) arithmetic speed, and (*d*) arithmetic power. It is required to find canonical variates for the two reading tests and the two arithmetic tests.

The correlations between the variates were—

	x_1	x_2	x_3	x_4
x_1	1·0000	0·6328	0·2412	0·0586
x_2	0·6328	1·0000	− 0·0553	0·0655
x_3	0·2412	− 0·0553	1·0000	0·4248
x_4	0·0586	0·0655	0·4248	1·0000

The determinant (28.83) becomes

$$\begin{vmatrix}
-\lambda & -0{\cdot}6328\lambda & 0{\cdot}2412 & 0{\cdot}0586 \\
-0{\cdot}6328\lambda & -\lambda & -0{\cdot}0553 & 0{\cdot}0655 \\
0{\cdot}2412 & -0{\cdot}0553 & -\lambda & -0{\cdot}4248\lambda \\
0{\cdot}0586 & 0{\cdot}0655 & -0{\cdot}4248\lambda & -\lambda
\end{vmatrix} = 0$$

or

$$0{\cdot}491{,}370\,\lambda^4 - 0{\cdot}078{,}803{,}4\,\lambda^2 + 0{\cdot}000{,}362{,}490 = 0,$$

giving $\lambda^2 = 0{\cdot}155{,}635$ or $0{\cdot}004{,}740$

with $\lambda = 0{\cdot}3945$ or $0{\cdot}0688$.

To find the transformed variates themselves we use (28.81). For instance, with the root $0{\cdot}3945$ for μ, we have

$$\begin{aligned} c^1 + 0{\cdot}6328\,c^2 - 0{\cdot}6114\,d^1 - 0{\cdot}1485\,d^2 &= 0 \\ 0{\cdot}6328\,c^1 + c^2 + 0{\cdot}1402\,d^1 - 0{\cdot}1660\,d^2 &= 0 \\ -0{\cdot}6114\,c^1 + 0{\cdot}1402\,c^2 + d^1 + 0{\cdot}4248\,d^2 &= 0 \\ -0{\cdot}1485\,c^1 - 0{\cdot}1660\,c^2 + 0{\cdot}4248\,d^1 + d^2 &= 0 \end{aligned}$$

The last equation is linearly dependent on the other three, so adds nothing. In the other three we solve for the ratios of c's and d's, finding

$$c^1 : c^2 : d^1 : d^2 = -2{\cdot}7772 : 2{\cdot}2655 : -2{\cdot}4404 : 1.$$

Thus the transformed variates are

$$\begin{aligned} k_1\,\xi^1 &= -2{\cdot}7772\,x_1 + 2{\cdot}2655\,x_2 \\ k_2\,\eta^1 &= -2{\cdot}4404\,x_3 + x_4, \end{aligned}$$

where k_1 and k_2 may be chosen so that the variances of ξ^1 and η^1 are unity, if desired. Similar equations with the root $0{\cdot}0688$ will give us a further pair of canonical co-ordinates. Those we have worked out have the maximum correlation, the other pair having the minimum and therefore being of less interest.

28.27. In practical cases it is of some importance to know whether an observed canonical correlation r_1, say, is significant of real correlation. The problem has been solved for large samples but not completely for small samples. We shall conclude this chapter with a short account of the main results which have been reached.

For large samples we shall show that, for the standard error of a canonical correlation,

$$\operatorname{var} r = \frac{1}{n}(1 - r^2)^2 \qquad (28.97)$$

a remarkable result showing that the variance is the same as for a product–moment coefficient.

Denoting as usual the sample covariance by a_{ij} we have to the first order

$$E\,(a_{ij}) = \sigma_{ij}. \qquad (28.98)$$

To the same order,

$$\Sigma\,(a_{ij}\,a_{kl}) = \frac{1}{n^2}\,E\left\{\sum_\alpha (x_{i\alpha}\,x_{j\alpha}) \sum_\beta (x_{k\beta}\,x_{l\beta})\right\}.$$

If $\alpha \neq \beta$ the sums on the right are independent, and there are $n\,(n-1)$ such cases. When $\alpha = \beta$ we have n terms such as

$$E\,(x_{i\alpha}\,x_{j\alpha}\,x_{k\alpha}\,x_{l\alpha}) = \sigma_{ij}\,\sigma_{kl} + \sigma_{il}\,\sigma_{jk} + \sigma_{ik}\,\sigma_{jl}, \qquad (28.99)$$

as follows from the consideration that the characteristic function of the multivariate normal form is

$$\exp\,(-\tfrac{1}{2}\sigma_{ij}\,t^i\,t^j)$$

(cf. **15.12**, vol. I, p. 376).

Hence we have

$$E(a_{ij}\,a_{kl}) = \frac{n(n-1)}{n^2}\sigma_{ij}\,\sigma_{kl} + \frac{n}{n^2}(\sigma_{ij}\,\sigma_{kl} + \sigma_{il}\,\sigma_{jk} + \sigma_{ik}\,\sigma_{jl})$$

$$= \sigma_{ij}\,\sigma_{kl} + \frac{1}{n}(\sigma_{il}\,\sigma_{jk} + \sigma_{ik}\,\sigma_{jl}). \quad . \quad . \quad . \quad . \quad . \quad (28.100)$$

Thus

$$E(da_{ij}\,da_{kl}) = E(a_{ij}\,a_{kl}) - \sigma_{ij}\,\sigma_{kl}$$

$$= \frac{1}{n}(\sigma_{il}\,\sigma_{jk} + \sigma_{ik}\,\sigma_{jl}). \quad . \quad . \quad . \quad . \quad . \quad (28.101)$$

Now for any canonical correlation r we have

$$\left.\begin{array}{ll} c^\alpha c^\beta a_{\alpha\beta} = 1, & d^i d^j a_{ij} = 1 \\ r = c^\alpha d^i a_{\alpha i} & \end{array}\right\}. \quad . \quad . \quad . \quad . \quad (28.102)$$

If now we define for the sampling deviations in c's and d's corresponding to deviations in the a's,

$$\Delta c^\alpha = \sum_{t,\,u} \frac{\partial c^\alpha}{\partial a_{tu}} \Delta a_{tu}, \quad . \quad . \quad . \quad . \quad . \quad (28.103)$$

we find

$$\left.\begin{array}{l} 2\,a_{\alpha\beta}\,c^\alpha\,\Delta c^\beta + c^\alpha\,c^\beta\,\Delta a_{\alpha\beta} = 0 \\ 2\,a_{ab}\,d^a\,\Delta d^b + d^a\,d^b\,\Delta a_{ab} = 0 \\ \Delta r_1 = a_{\alpha b}\,c^\alpha\,\Delta d^b + a_{\alpha b}\,d^b\,\Delta c^\alpha + c^\alpha\,d^b\,\Delta a_{\alpha b} \end{array}\right\}. \quad . \quad . \quad . \quad (28.104)$$

Without loss of generality we may now suppose the variates canonical and hence put $c^1 = 1,\ c^2 = c^3 = \ldots = c^p = 0,\ d^1 = 1,\ d^2 = \ldots = d^q = 0.$ We then find—

$$\left.\begin{array}{ll} 2\Delta c^1 + \Delta a_{11} = 0, & 2\Delta d^1 + \Delta a_{p+1,\,p+1} = 0 \\ \Delta r_1 = r_1\,\Delta d^1 + r_1\,\Delta c^1 + \Delta a_{1,\,p+1} & \end{array}\right\}. \quad . \quad . \quad (28.105)$$

Substituting from the first two in the third of these equations we have

$$\Delta r_1 = \Delta a_{1,\,p+1} - \tfrac{1}{2}r_1(\Delta a_{11} + \Delta a_{p+1,\,p+1}). \quad . \quad . \quad . \quad (28.106)$$

Similar equations apply for any other simple root, e.g.

$$\Delta r_2 = \Delta a_{2,\,p+2} - \tfrac{1}{2}r_2(\Delta a_{22} + \Delta a_{p+2,\,p+2}).$$

Squaring these equations and substituting from (28.101) we find

$$nE(\Delta r_1)^2 = (1 - r_1^2)^2$$
$$E(\Delta r_1,\,\Delta r_2) = 0.$$

It follows that

$$\left.\begin{array}{r} \operatorname{var} r_1 = \dfrac{1}{n}(1 - \rho_1^2)^2 \\ \operatorname{cov}(r_1,\,r_2) = 0 \end{array}\right\} \quad . \quad . \quad . \quad . \quad (28.107)$$

to our order of approximation.

28.28. Equation (28.107) applies to a simple non-vanishing correlation. If a canonical correlation vanishes and $p = q$, the result holds, with the qualification that sample values of r near the zero root must be allowed to have positive or negative values, or alternatively that the distribution of r is that of *absolute* values of a normal variate (cf. Exercise 28.7). If $p = 2,\ q > 2$ a zero root is of multiplicity q at least. In this case, if it has exactly

multiplicity q, nr^2 is distributed as χ^2 with $q - 1$ degrees of freedom. For the proof of this result see Hotelling (1936*b*).

There is another rather curious difficulty in testing the significance of roots of the equation giving the canonical correlations, namely, that if several roots exist it is not possible to relate them with certainty to specified parent correlations—any one might have arisen from any one of the parent values. This is not serious for large samples when the roots are distinct, since the sample values cluster closely round the parent values; but for small samples or canonical correlations in the parent which are close together it presents a theoretical problem of a novel kind. See Hotelling (1936*b*) and Bartlett (1941) on this point.

28.29. We proceed to find the sampling distribution of canonical correlations in the case when the parent values are all zero and the p-variates and q-variates accordingly independent.

Reverting to equation (28.87) in the form appropriate to samples, we have

$$| \lambda^2 a_{\beta\gamma} - a_{i\beta}\, a^{ik}\, a_{\gamma k} | = 0. \qquad (28.108)$$

We write

$$t_{\beta\gamma} = a_{i\beta}\, a^{ik}\, a_{\gamma k} \qquad (28.109)$$

and

$$a_{\beta\gamma} = z_{\beta\gamma} + t_{\beta\gamma}, \qquad (28.110)$$

so that (28.108) becomes

$$| \lambda^2 (z_{\beta\gamma} + t_{\beta\gamma}) - t_{\beta\gamma} | = 0. \qquad (28.111)$$

The significance of this device is that z and t are distributed independently in Wishart's form, as we now proceed to show.

One instructive way of looking at the problem is to consider the regression of the p-way vector y on a q-way vector x. Corresponding to the univariate equation

$$y = bx + e, \qquad (28.112)$$

where e is a residual, we have

$$y_\alpha = b^i_\alpha\, x_i + x_{i\alpha}, \qquad (28.113)$$

where the b's are given by minimising the sum of n values

$$\Sigma\, (y_\alpha - b^i_\alpha\, x_i)^2$$

namely, by

$$\Sigma\, (y_\alpha\, x_i) - b^k_\alpha\, \Sigma\, (x_k\, x_i) = 0$$

or, in our notation for canonical variates,

$$a_{\alpha i} - b^k_\alpha\, a_{ki} = 0,$$

which yields

$$b^k_\alpha = a_{\alpha i}\, a^{ki}. \qquad (28.114)$$

We may analyse the variance of y in the form—

$$\begin{aligned} \Sigma\, (y^2_\alpha) &= \Sigma\, (b^i_\alpha\, x_i + x_{i\alpha})^2 \\ &= b^i_\alpha\, b^k_\alpha\, a_{ik} + \Sigma\, (x_{i\alpha})^2, \end{aligned} \qquad (28.115)$$

corresponding to the univariate case

$$\Sigma\, (y^2) = b^2\, \Sigma\, (x^2) + \Sigma\, (e^2),$$

and the two constituents on the right in (28.115) are independent, just as in the univariate case. This may be shown by a direct extension of **22.19**.

Furthermore, if we wish to find the linear function of the y's, say $\lambda^\alpha y_\alpha$, which has maximum correlation with the x's, we have to maximise the ratio

$$\frac{\Sigma\,(\lambda^\alpha b^i_\alpha x_i)^2}{\Sigma\,(\lambda^\alpha y_\alpha)^2} = \frac{\lambda^\alpha \lambda^\beta b^i_\alpha b^j_\beta a_{ij}}{\lambda^\alpha \lambda^\beta a_{\alpha\beta}} = r^2. \qquad (28.116)$$

This is equivalent to maximising unconditionally

$$\lambda^\alpha \lambda^\beta (b^i_\alpha b^j_\beta a_{ij} - r^2 a_{\alpha\beta}) = 0,$$

giving, for r^2, the equation—

$$| \, b^i_\alpha b^j_\beta a_{ij} - r^2 a_{\alpha\beta} \, | = 0. \qquad (28.117)$$

Now in virtue of (28.114) this reduces to

$$| \, r^2 a_{\alpha\beta} - a_{ij} a_{\alpha m} a^{mi} a_{\beta p} a^{pj} \, | = 0$$

or

$$| \, r^2 a_{\alpha\beta} - a_{\alpha j} a^{pj} a_{\beta p} \, | = 0, \qquad (28.118)$$

which is equivalent to (28.108) with a slight change of notation. This must be so, for we arrived at both equations on essentially the same assumptions. Now we see that the term on the right in the determinant of (28.118) is the first item on the right of the variance analysis given by (28.115), and the other term in the determinant is the sum $\Sigma\,(y^2)$ of the analysis. It follows that z and t of (28.111) are independent, for they are the constituent items of the analysis. Furthermore, the z's will be distributed as sums of squares or products about the means with $n - q$ degrees of freedom, that is in Wishart's form; and similarly the t's are distributed as q sums of squares or products about the origin, i.e. in Wishart's form with $n = q + 1$.

28.30. Without loss of generality we may take the parent variances to be unity; the covariances are zero by hypothesis. The joint distribution of z and t is then, from (28.26),

$$dF = \frac{| \, t \, |^{\frac{1}{2}(q-p-1)} \, | \, z \, |^{\frac{1}{2}(n-q-p-2)} \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{p} (t_{ii} + z_{ii}) \right\} \Pi \, dt \, dz}{2^{\frac{1}{2}p(n+1)} \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^{p} \left\{ \Gamma\left(\frac{q+1-i}{2}\right) \Gamma\left(\frac{n-q-i}{2}\right) \right\}}. \qquad (28.119)$$

In the determinant

$$| \, \lambda^2 (z + t) - t \, | = 0$$

put $u = \lambda^2$ and let the roots in u be arranged in descending order of magnitude. Consider the distribution for a given value of t_{ij} and z_{ij} which in particular we take to be δ_{ij}. Let us choose new variates from a set ξ_{jk} obeying the orthogonality conditions—

$$\begin{aligned} \sum_{k=1}^{p} (\xi_{ik} \xi_{jk}) &= \delta_{ij} \\ &= 0 \text{ if } i \neq j \\ &= 1 \text{ if } i = j. \end{aligned} \qquad (28.120)$$

Make the transformation

$$t_{ij} = \sum_k (\xi_{ik} \xi_{jk} u_k) \qquad (28.121)$$

$$t_{ij} + z_{ij} = \sum_k (\xi_{ik} \xi_{jk}) = \boldsymbol{\delta}_{ij}. \qquad (28.122)$$

Instead of the $\frac{1}{2}p(p+1)$ values of t_{ij} we will take the p values of u and $\frac{1}{2}p(p-1)$ of the ξ's as our new variates. We have

$$|t| = |\xi_{ik}\xi_{jk}u_k| = \prod_{k=1}^{p} u_k \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.123)$$

$$|z| = |\xi_{ik}\xi_{jk}(1-\delta_{ij}u_k)| = \prod_{k=1}^{p}(1-u_k) \quad . \quad . \quad . \quad (28.124)$$

and have only to consider the Jacobian. This is clearly of degree $\frac{1}{2}p(p-1)$ in u, for the Jacobian of t and $z+t$ is the same as that of t and z and only t contributes factors in u in the former. Furthermore, every term $(u_i - u_j)$, $i<j$ is a factor of J. For consider $u_1 - u_2$ and let us take as our ξ-variates those for which $j>i$. Then to satisfy the conditions on the others, derivable from (28.120),

$$\frac{\partial}{\partial\xi_{ij}}\sum_k(\xi_{ik}\xi_{jk}) = 0,$$

we must have

$$\frac{\partial\xi_{i1}}{\partial\xi_{12}} = -\frac{\xi_{i2}}{\xi_{11}}, \qquad \frac{\partial\xi_{i2}}{\partial\xi_{12}} = \frac{\xi_{i1}}{\xi_{11}}$$

$$\frac{\partial\xi_{ij}}{\partial\xi_{12}} = 0, \qquad j>2,$$

whence

$$\frac{\partial t_{ij}}{\partial\xi_{12}} = \frac{\partial}{\partial\xi_{12}}\Sigma(\xi_{ik}\xi_{jk}u_k)$$

$$= -\frac{\xi_{i2}\xi_{j1}}{\xi_{11}}(u_1-u_2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (28.125)$$

Thus every term $(u_i - u_j)$ occurs in J, and there can be no further factors in u because the power in u is $\frac{1}{2}p(p-1)$.

Substituting in (28.119) we have, integrating out the ξ-variates,

$$dF = c\prod_{i=1}^{p}\{u_i^{\frac{1}{2}(q-p-1)}(1-u_i)^{\frac{1}{2}(n-q-p-2)}\}\,\Pi(u_i-u_j)\,\Pi\,du \quad . \quad (28.126)$$

where

$$c = \frac{k}{\Pi\left\{\Gamma\left(\frac{q+1-i}{2}\right)\Gamma\left(\frac{n-q-i}{2}\right)\right\}} \quad . \quad . \quad . \quad (28.127)$$

The constant k arises from terms involving n and p in the original density and from the Jacobian. It therefore does not involve q and may be written $k(n,p)$. Evaluation of k by direct integration is a matter of some difficulty, but we may find it indirectly as follows :—

In (28.126), if we increase q and n by $2s$, the corresponding value of c is

$$\frac{k(n+2s,p)}{\Pi\left\{\Gamma\left(\frac{q+1+2s-i}{2}\right)\Gamma\left(\frac{n-q-i}{2}\right)\right\}} \quad . \quad . \quad . \quad . \quad (28.128)$$

The only other term in (28.126) which is affected is that in $\Pi(u)$ and, with the original

c of (28.127), the integral of the distribution so modified would give us the moment of order s of $\Pi(u)$, namely of $|t|$. This may be found in the manner of **28.15** to be

$$\Pi \frac{\Gamma\left(\frac{q+1+2s-i}{2}\right)\Gamma\left(\frac{n+1-i}{2}\right)}{\Gamma\left(\frac{q+1-i}{2}\right)\Gamma\left(\frac{n+2s+1-i}{2}\right)} \quad . \quad . \quad . \quad (28.129)$$

(see Exercise **28.11**). It follows that

$$\frac{k(n+2s, p)}{k(n, p)} = \Pi \frac{\Gamma\left(\frac{n+2s-i}{2}\right)}{\Gamma\left(\frac{n-i}{2}\right)}, \quad . \quad . \quad . \quad . \quad (28.130)$$

whence

$$k(n, p) = \Pi\, \Gamma\left(\frac{n-i}{2}\right) f(p). \quad . \quad . \quad . \quad . \quad (28.131)$$

It remains to evaluate $f(p)$. To do so we make the substitution in (28.126)

$$u_i = \frac{2v_i}{n},$$

letting n tend to infinity. Our distribution becomes

$$dF = \frac{f(p)\,(\Pi v)^{\frac{1}{2}(q-p-1)}}{\Gamma\left(\frac{p+2-i}{2}\right)} \exp(-\Sigma v_i)\, \Pi(v_i - v_j)\, \Pi\, dv. \quad . \quad . \quad (28.132)$$

This may be reduced by successive substitutions of the type

$$v_1 = w_1, \qquad v_j = w_j + v_1, \qquad j > 1,$$

and choosing q at each stage so that the term in $\Pi(v)$ vanishes (as we may, since the result is independent of q). On integration for v_1, then repeating the process, and so on, we find

$$\frac{f(p)}{\Pi\, \Gamma\left(\frac{p+2-i}{2}\right)} \; \frac{\Pi\, \Gamma(p+1-i)}{2^{\frac{1}{2}p(p-1)}} = 1.$$

Using the relation

$$\Gamma(x)\,\Gamma(x+\tfrac{1}{2}) = 2^{-2x+1}\sqrt{\pi}\,\Gamma(2x),$$

we have

$$f(p) = \frac{\pi^{\frac{1}{2}p}}{\Pi\, \Gamma\left(\frac{p+1-i}{2}\right)}. \quad . \quad . \quad . \quad . \quad (28.133)$$

Thus our distribution is finally

$$dF = c\, \Pi\, \{u^{\frac{1}{2}(q-p-1)}(1-u)^{\frac{1}{2}(n-p-q-2)}\}\, \Pi(u_i - u_j)\, \Pi\, du, \quad . \quad (28.134)$$

where

$$c = \pi^{\frac{1}{2}p} \prod_{i=1}^{p} \frac{\Gamma\left(\frac{n-i}{2}\right)}{\Gamma\left(\frac{q+1-i}{2}\right)\Gamma\left(\frac{p+1-i}{2}\right)\Gamma\left(\frac{n-q-i}{2}\right)}, \quad . \quad (28.135)$$

a remarkable form obtained in the general case by Fisher (1939*b*), P. L. Hsu (1939*b*), and Roy (1939*b*).

We have supposed throughout that $q \geqslant p$. In the contrary case we reverse the roles of q and p and hence merely have to interchange p and q in (28.134) and (28.135).

28.31. Let us consider some special cases. When $q = 1$ the distribution becomes

$$dF = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{p}{2}\right)} u_1^{\frac{1}{2}(p-2)}(1-u_1)^{\frac{1}{2}(n-p-3)}\,du_1, \quad . \quad (28.136)$$

confirming the distribution of equation (28.40) leading to Hotelling's distribution; for the canonical correlation is then the multiple correlation between the q-variate and the p-variates; and as the former is measured from its mean there is one fewer degree of freedom, i.e. n is replaced by $n - 1$.

When $q = 2$ we have

$$dF = \frac{\pi^{\frac{1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{n-p-2}{2}\right)\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}(u_1u_2)^{\frac{1}{2}(p-3)}\{(1-u_1)(1-u_2)\}^{\frac{1}{2}(n-p-4)}$$
$$\times (u_1 - u_2)\,du_1\,du_2. \quad . \quad (28.137)$$

Writing

$$(1-u_1)(1-u_2) = v,$$
$$u_1 + u_2 = w,$$

we find

$$dF = \frac{\Gamma(n-2)}{4\Gamma(n-p-2)\,\Gamma(p-1)}(v-1+w)^{\frac{1}{2}(p-3)}\,v^{\frac{1}{2}(n-p-4)}\,dv\,dw. \quad . \quad (28.138)$$

For given v the limits of w are $1 - v$ and $2(1-\sqrt{v})$, and integrating for w we find

$$dF = \frac{\Gamma(n-2)}{4\Gamma(n-p-2)\,\Gamma(p-1)} \cdot \frac{2}{p-1}(1-\sqrt{v})^{p-1}(\sqrt{v})^{n-p-4}\,dv$$

or, for $\sqrt{v}$,

$$dF = \frac{1}{B(n-p-2,\,p)}(1-\sqrt{v})^{p-1}(\sqrt{v})^{n-p-3}\,d\sqrt{v}, \quad . \quad . \quad . \quad (28.139)$$

a result due to Wilks—cf. equation (28.62).

28.32. The distribution of the u's does not immediately provide a test of significance of the canonical correlations, except when there is only one of them. The criterion

$$v = \Pi(1-u) \quad . \quad . \quad . \quad . \quad . \quad (28.140)$$

is sometimes useful in the general case for testing simultaneously the departure of the u's from zero. Cf. Exercises 28.11 and 28.12.

NOTES AND REFERENCES

Among earlier papers in which various aspects of the multivariate problem began to be studied, reference may be made to Karl Pearson (1926*b*) on the "coefficient of racial likeness" and Ragnar Frisch (1929), who independently arrived at the dispersion matrix and proposed to call its determinant in standard measure the "scatterance". Reference

to the papers by Wishart (1928), Wishart and Bartlett (1933*c*) and Hotelling (1931) on the generalised product-moment distribution and the generalised "Student" ratio has been made in the text.

In more recent literature three lines of development are discernible :—

(*a*) American writers have developed the theory of canonical correlation and multiple analysis mainly on algebraic and analytical lines. See Hotelling (1933, 1936*b*), Wilks (1932*e*, 1934, 1935*b*, 1935*c*, 1936, 1943), Girshik (1939), and Madow (1938).

(*b*) English schools have investigated the theory of discriminant functions and developed the sampling theory of canonical roots. See R. A. Fisher (1936*a*, *b*, 1938*c*, 1939*b*, 1940*d*), P. L. Hsu (1938*c*, 1939*b*, 1941*a*, *c*, *d*), and for illustrative material Martin (1936), Barnard (1935), Fairfield Smith (1936) and Wallace and Travers (1938). See also Bartlett (1934*b*, 1938*c*, 1939*b*, *c*, 1941), E. S. Pearson and Wilks (1933*b*), Welch (1939*b*), Lawley (1938) and Bishop (1939). Simaika (1941) has proved that tests based on Hotelling's T and the multiple correlation coefficient are uniformly most powerful in the class depending on a single parameter.

(*c*) The Indian school, whose contribution has not been referred to in this chapter, has developed some interesting work based on what is known as the D^2-statistic. See Mahalanobis (1930, 1936*a*), Mahalanobis, Bose and Roy (1936*b*), R. C. Bose (1936*a*), R. C. Bose and Roy (1938*c*), and later papers in *Sankhyā*. If, with two samples from p-variate populations, d_i is the difference of sample means for the ith variate, the studentised D^2-statistic is

$$D^2 = \frac{1}{p} a^{ij} d_i d_j,$$

where a^{ij} refers to the reciprocal of the sample dispersion matrix. Bose and Roy have shown that in normal samples this has the same distribution as one of Fisher's forms for the multiple correlation coefficient. The corresponding parameter for the population

$$\Delta^2 = \frac{1}{p} \alpha^{ij} \delta_i \delta_j$$

is known as Mahalanobis's generalised distance.

EXERCISES

28.1. In a four-variate normal distribution show that the correlation between the covariances a_{12} and a_{34} is

$$\frac{\rho_{13}\rho_{24} + \rho_{14}\rho_{23}}{\{(1+\rho_{12}^2)(1+\rho_{34}^2)\}^{\frac{1}{2}}}$$

(Wishart, 1928.)

28.2. For a pair of normal variates with correlation ρ, show that, defining v by

$$v = \frac{n\, a_{12}}{\sigma_1 \sigma_2 (1-\rho^2)},$$

we have for the frequency function of v

$$f(v) = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}\, e^{\rho v}}{\sqrt{\pi}\, 2^{\frac{1}{2}n-1}\, \Gamma\left(\dfrac{n-1}{2}\right)} \{v^{\frac{1}{2}n-1} K_{\frac{1}{2}n-1}(v)\}$$

for $v > 0$ and a similar expression with $-v$ for v inside curly brackets if $v < 0$. Here K is the Bessel function of second kind with imaginary argument.

(Wishart and Bartlett, 1933*c*. See also K. Pearson and others, 1929.)

28.3. Show that if k sets of variates $a_{ij}^{(h)}$, $h = 1 \ldots k$; $i, j = 1 \ldots p$ are each distributed in Wishart's form, with sample numbers $n_1 \ldots n_k$, then the variates

$$a_{ij} = \sum_{h=1}^{k} a_{ij}^{(h)}$$

are also distributed in Wishart's form with $n = \sum_{h=1}^{k} (n_h)$. (This follows readily from the characteristic function. It is a generalisation of the additive properties of χ^2.)

28.4. If a sample of n is chosen from a p-variate normal population, the variates being grouped into k classes $x_1, x_2 \ldots x_{p_1}$; $x_{p_1+1} \ldots x_{p_1+p_2}$; $\ldots$; $x_{p_1+\ldots p_{k-1}+1} \ldots x_p$, consider the function—

$$W = \frac{| r_{ij} |}{| r_{ij}^{(0)} |}$$

where $r_{ii} = 1$ and $r_{ij}^{(0)}$ is zero if the variates belong to different classes and equals the correlation r_{ij} if they belong to the same class.

By considering the function

$$\lambda = W^{\frac{1}{2}n}$$

show that

$$\mu_r'(W) = \prod_{t=1}^{k} \prod_{i=1}^{p_t} \left\{ \frac{\Gamma\left(\frac{n-i}{2}\right)}{\Gamma\left(\frac{n-i}{2}+r\right)} \right\} \prod_{i=1}^{p} \frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}.$$

(Wilks, 1935*b*. The distribution provides a test of the independence of k sets of normal variates.)

28.5. As a particular case of the last exercise, show that if a single variate x_1 is independent of a second set $x_2 \ldots x_p$, then—

$$\mu_r(W) = \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-p}{2}+r\right)}{\Gamma\left(\frac{n-1}{2}+r\right)\Gamma\left(\frac{n-p}{2}\right)};$$

and hence find the distribution of the multiple correlation coefficient when the parent coefficient is zero.

(Wilks, 1935*b*.)

28.6. Show algebraically that Hotelling's T is invariant under linear transformations of the p variates.

28.7. If the determinantal equation (28.83) with $p = q$ has a double root equal to zero, show that for large samples the value of r corresponding to the canonical correlation

is given by omitting all terms in the determinant when expanded, except those in λ^2 and λ^0. Noting that the latter is a perfect square, show that r is the ratio of a polynomial in the sample dispersions to a non-vanishing function regular in the neighbourhood of zero. Hence that (28.107) holds when $\rho = 0$.

(Hotelling, 1936*b*.)

28.8. In the notation of **28.23**, if

$$A = |\sigma_{\alpha\beta}|, \qquad B = |\sigma_{ij}|$$

$$C = \left|\begin{array}{c:c} 0 & \sigma_{\alpha i} \\ \hdashline \sigma_{i\alpha} & \sigma_{ij} \end{array}\right|, \qquad D = \left|\begin{array}{c:c} \sigma_{\alpha\beta} & \sigma_{\alpha i} \\ \hdashline \sigma_{i\alpha} & \sigma_{ij} \end{array}\right|$$

show that the *vector correlation coefficient* K defined by

$$K^2 = \frac{(-1)^p C}{AB}$$

and the square of the *vector alienation coefficient* Z defined by

$$Z = \frac{D}{AB}$$

are invariant under linear transformations of the variate. Also that

$$K = \pm \rho_1 \rho_2 \dots \rho_p$$
$$Z = (1 - \rho_1^2)(1 - \rho_2^2) \dots (1 - \rho_p^2)$$

where the ρ's are canonical correlations.

(Hotelling, 1936*b*.)

28.9. In the notation of the previous exercise, k and z being the sample values of K and Z, show that if the population canonical correlations are all distinct,

$$\text{var } k = \frac{1}{n} K^2 \sum_{i=1}^{p} \left\{ \frac{(1 - \rho_i^2)^2}{\rho_i^2} \right\}$$

$$\text{var } z = \frac{4}{n} Z^2 \sum_{i=1}^{p} \rho_i^2$$

$$\text{cov}(k, z) = -\frac{2}{n} KZ \sum_{i=1}^{p} (1 - \rho_i^2).$$

In particular, when $p = 2$,

$$\text{var } k = \frac{1}{n} \{ (1 - K^2)^2 - Z(1 + K^2) \}$$

$$\text{var } z = \frac{4Z^2}{n} (1 - Z + K^2)$$

$$\text{cov}(k, z) = -\frac{2}{n} KZ (1 + Z - K^2).$$

(Hotelling, 1936*b*.)

28.10. In the previous exercise, with $p = q = 2$, show that, in standard measure,

$$k = \frac{r_{13}\,r_{24} - r_{14}\,r_{23}}{\{(1 - r_{12}^2)(1 - r_{34}^2)\}^{\frac{1}{2}}}$$

and hence derive a test of significance of the "tetrad difference" $r_{13}\,r_{24} - r_{14}\,r_{23}$.

(Hotelling, 1936*b*.)

28.11. In the notation of Exercise 28.9, show that

$$E(k^{\alpha} z^{\beta}) = \prod_{i=1}^{p} \left[\frac{\Gamma\left(\frac{q + \alpha + 1 - i}{2}\right) \Gamma\left(\frac{n - q + 2\beta - i}{2}\right) \Gamma\left(\frac{n - i}{2}\right)}{\Gamma\left(\frac{q + 1 - i}{2}\right) \Gamma\left(\frac{n - q - i}{2}\right) \Gamma\left(\frac{n + \alpha + 2\beta - i}{2}\right)} \right].$$

(Girshik, 1939.)

28.12. Find the characteristic function of $-\log z$, where z is defined as in the previous exercise, and hence show that $-n \log z$ or, to a better approximation, $-\{n - 1 - \frac{1}{2}(p + q + 1)\}\log z$ tends to be distributed as χ^2 with pq degrees of freedom when n is large.

(Bartlett, 1938*c*.)

CHAPTER 29

TIME-SERIES—(1)

29.1. A time-series, as its name indicates, is a series of values assumed by a variable at different points of time. We shall consider only cases where the variable is univariate and shall denote its value at time t by u_t. The study of such series forms an important branch of statistics because the majority of types of time-variation encountered in practice are not of the regular functional type in which u_t can be represented exactly by a mathematical function of t, but present in some degree those irregularities of a random character which can only be discussed in terms of probability. One of our main problems, in fact, will be to isolate systematic from casual effects in the series so as to be able to study them separately.

29.2. In general it is possible to observe a time-variable at any instant, and thus the temporal intervals between successive members of the series need not be the same. Practice and theory alike, however, usually require the observations to occur at regular intervals, and in the sequel we shall assume, unless the contrary is specifically stated, that the interval from each observation to the next is the same throughout the series. As a matter of convenience we may take this interval as our time-unit and write the series as

$$u_1, u_2, u_3, \ldots u_t, \ldots \qquad (29.1)$$

where t must be an integer. Where a series extends backwards and forwards from some given point which we wish to regard as origin we may write it as

$$\ldots u_{-t}, \ldots u_{-2}, u_{-1}, u_0, u_1, u_2, \ldots u_t, \ldots \qquad (29.2)$$

In this chapter and the next we shall study the way in which u_t varies with t, such variation being in general of the stochastic type, that is to say, involving random variables.

Some Examples of Time-series

29.3. Tables 29.1 to 29.5 provide some examples of the kind of variation encountered in practice. Table 29.1 (illustrated in Fig. 29.1) gives the annual yields per acre of barley in England and Wales from 1884 to 1939. Table 29.2 (Fig. 29.2) shows the human population of England and Wales at ten-yearly intervals from 1811 to 1931. Table 29.3 (Fig. 29.3) gives the sheep population of England and Wales for each year from 1867 to 1939. Table 29.4 (Fig. 29.4) gives the annual rainfall in London for each year from 1813 to 1912. Table 29.5 (Fig. 29.5) gives the average egg-production per laying hen in the U.S.A. for each month of the years 1938 to 1940.

TABLE 29.1

Annual Yields per Acre of Barley in England and Wales from 1884 to 1939.

(Data from the *Agricultural Statistics.*)

Year.	Yield per acre (cwts.).	Year.	Yield per acre (cwts.).	Year.	Yield per acre (cwts.).	Year.	Yield per acre (cwts.).
1884	15·2	1898	16·9	1912	14·2	1926	16·0
85	16·9	99	16·4	13	15·8	27	16·4
86	15·3	1900	14·9	14	15·7	28	17·2
87	14·9	01	14·5	15	14·1	29	17·8
88	15·7	02	16·6	16	14·8	30	14·4
89	15·1	03	15·1	17	14·4	31	15·0
90	16·7	04	14·6	18	15·6	32	16·0
91	16·3	05	16·0	19	13·9	33	16·8
92	16·5	06	16·8	20	14·7	34	16·9
93	13·3	07	16·8	21	14·3	35	16·6
94	16·5	08	15·5	22	14·0	36	16·2
95	15·0	09	17·3	23	14·5	37	14·0
96	15·9	10	15·5	24	15·4	38	18·1
97	15·5	11	15·5	25	15·3	39	17·5

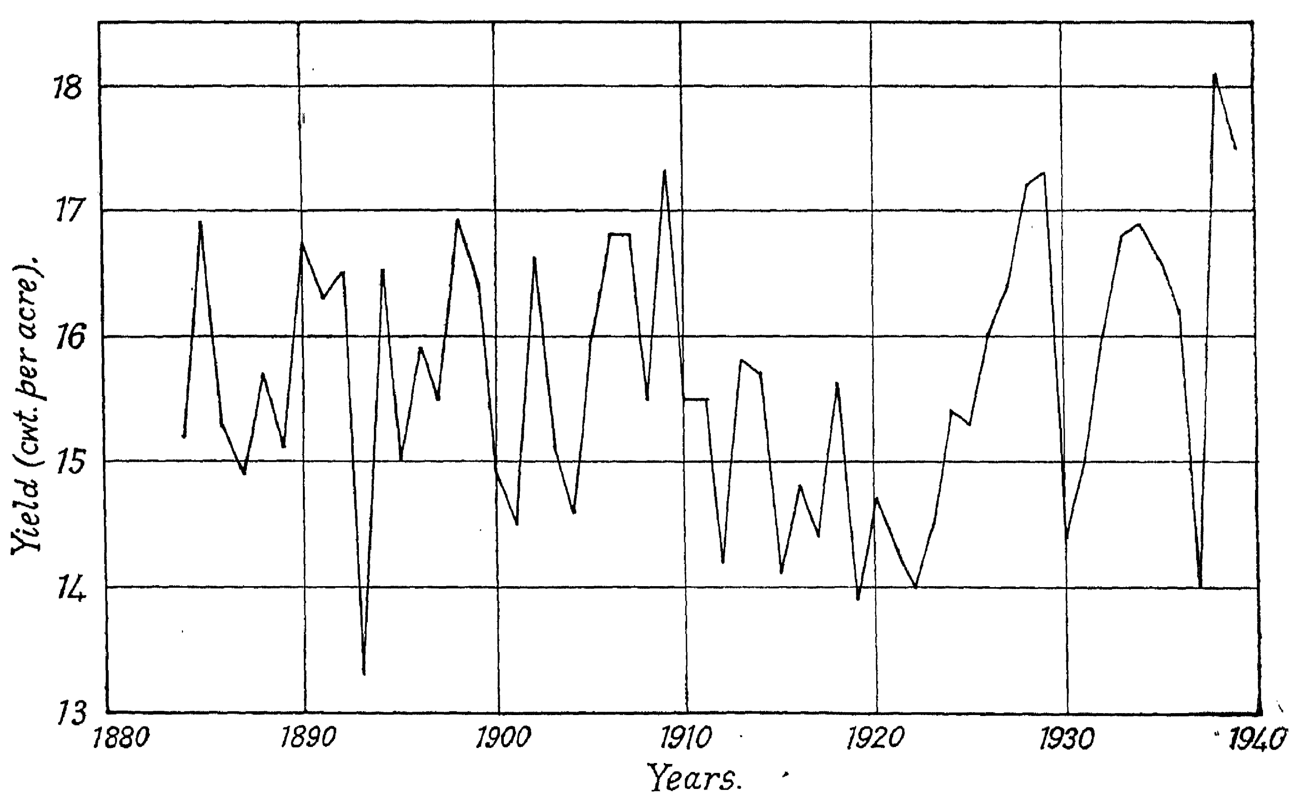


FIG. 29.1.—Graph of the Data of Table 29.1 (Barley Yields per Acre).

TABLE 29.2

Population of England and Wales at Ten-Yearly Intervals from 1811 to 1931.

(Data from the Registrar-General's *Statistical Review*, 1933, Part II.)

Year.	Population (millions).
1811	10·16
21	12·00
31	13·90
41	15·91
51	17·93
61	20·07
71	22·71
81	25·97
91	29·00
1901	32·53
11	36·07
21	37·89
31	39·95

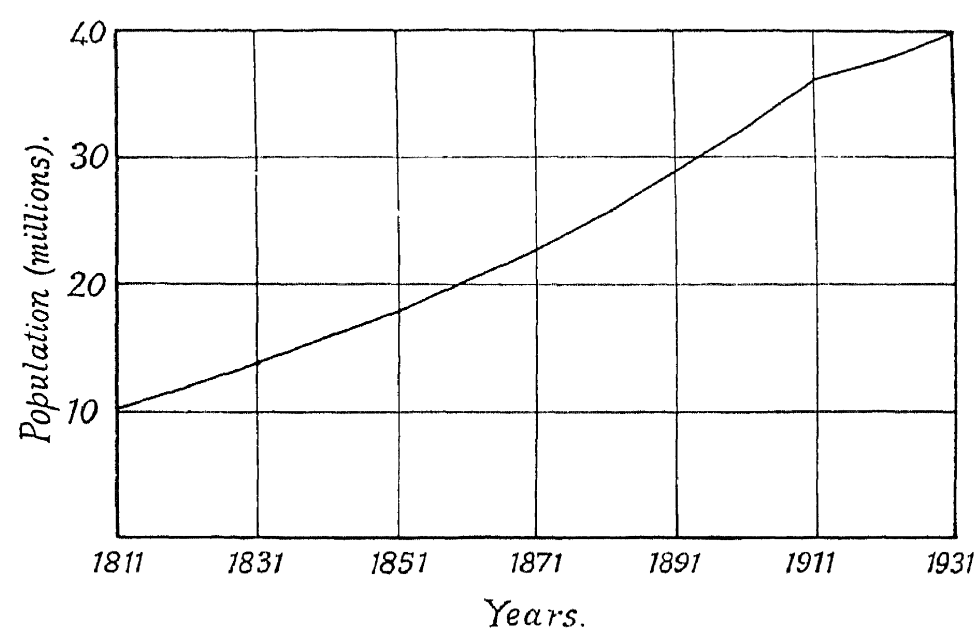


FIG. 29.2.—Graph of the Data of Table 29.2 (Population of England and Wales).

TABLE 29.3

Sheep Population of England and Wales for each Year from 1867 to 1939.

(Data from the *Agricultural Statistics.*)

Year.	Population (10,000).	Year.	Population (10,000).	Year.	Population (10,000).	Year.	Population (10,000).
1867	2203	1886	1892	1905	1823	1924	1484
68	2360	87	1919	06	1843	25	1597
69	2254	88	1853	07	1880	26	1686
70	2165	89	1868	08	1968	27	1707
71	2024	90	1991	09	2029	28	1640
72	2078	91	2111	10	1996	29	1611
73	2214	92	2119	11	1933	30	1632
74	2292	93	1991	12	1805	31	1775
75	2207	94	1859	13	1713	32	1850
76	2119	95	1856	14	1726	33	1809
77	2119	96	1924	15	1752	34	1653
78	2137	97	1892	16	1795	35	1648
79	2132	98	1916	17	1717	36	1665
80	1955	99	1968	18	1648	37	1627
81	1785	1900	1928	19	1512	38	1791
82	1747	01	1898	20	1338	39	1797
83	1818	02	1850	21	1383		
84	1909	03	1841	22	1344		
85	1958	04	1824	23	1384		

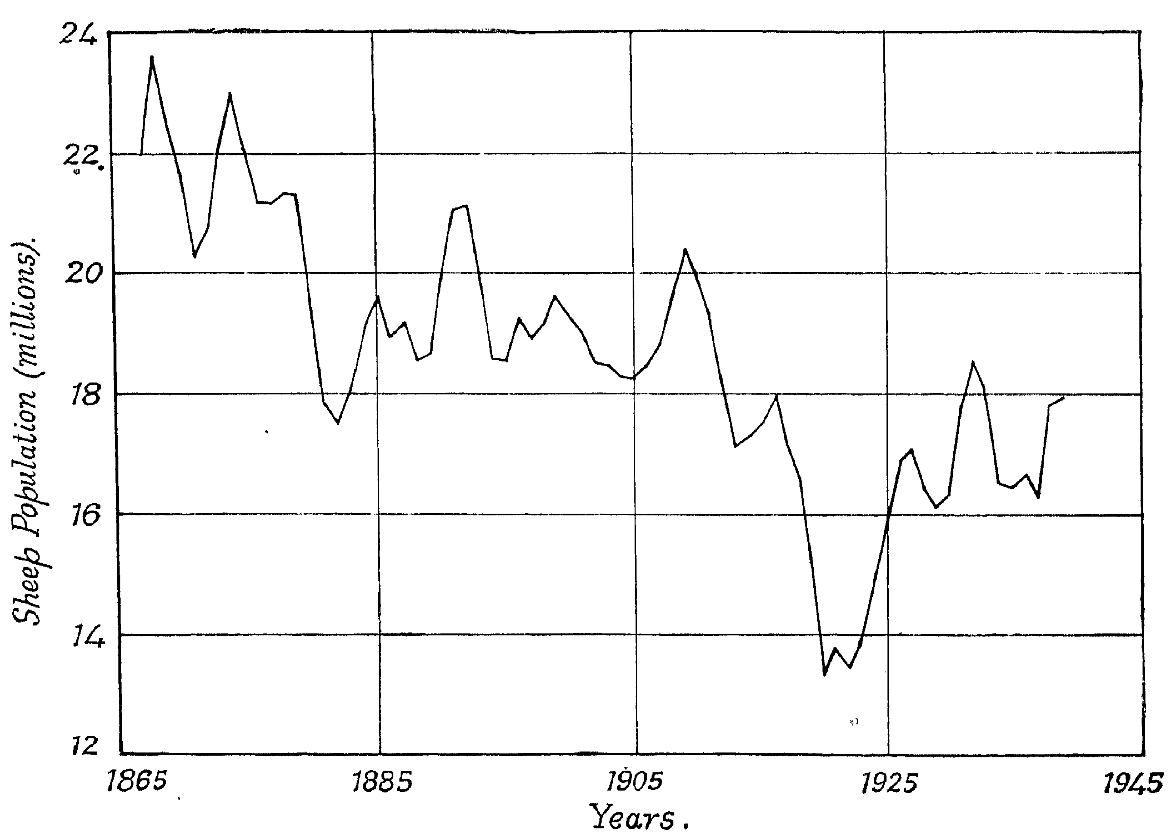


FIG. 29.3.—Graph of the Data of Table 29.3 (Sheep Population).

TABLE 29.4

Total Annual Rainfall at London in Inches, for each Year from 1813 to 1912.

(Data from D. Brunt, *Phil. Trans.* A, **225**, 247, 1925.)

Year.	Rainfall (inches).	Year.	Rainfall (inches).	Year.	Rainfall (inches).	Year.	Rainfall (inches).
1813	23·56	1838	21·63	1863	21·59	1888	27·74
14	26·07	39	27·49	64	16·93	89	23·85
15	21·86	40	19·43	65	29·48	90	21·23
16	31·24	41	31·13	66	31·60	91	28·15
17	23·65	42	23·09	67	26·25	92	22·61
18	23·88	43	25·85	68	23·40	93	19·80
19	26·41	44	22·65	69	25·42	94	27·94
20	22·67	45	22·75	70	21·32	95	21·47
21	31·69	46	26·36	71	25·02	96	23·52
22	23·86	47	17·70	72	33·86	97	22·86
23	24·11	48	29·81	73	22·67	98	17·69
24	32·43	49	22·93	74	18·82	99	22·54
25	23·26	50	19·22	75	28·44	1900	23·28
26	22·57	51	20·63	76	26·16	01	22·17
27	23·00	52	35·34	77	28·17	02	20·84
28	27·88	53	25·89	78	34·08	03	38·10
29	25·32	54	18·65	79	33·82	04	20·65
30	25·08	55	23·06	80	30·28	05	22·97
31	27·76	56	22·21	81	27·92	06	24·26
32	19·82	57	22·18	82	27·14	07	23·01
33	24·78	58	18·77	83	24·40	08	23·67
34	20·12	59	28·21	84	20·35	09	26·75
35	24·34	60	32·24	85	26·64	10	25·36
36	27·42	61	22·27	86	27·01	11	24·79
37	19·44	62	27·57	87	19·21	12	27·88

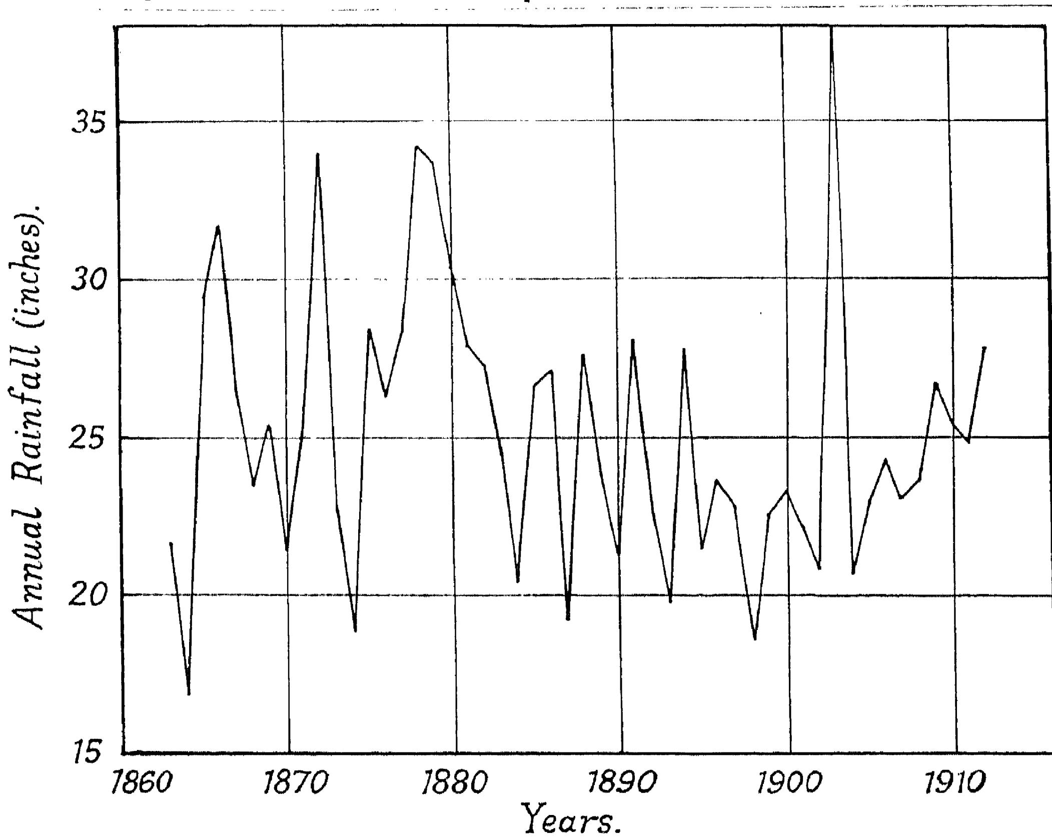


FIG. 29.4.—Graph of the Last 50 Terms of the Data of Table 29.4 (Rainfall).

TABLE 29.5

Average Number of Eggs per Laying Hen in the U.S.A. for each Month of the Years 1938–1940.

(Data from Report of the Bureau of Agricultural Economics, U.S. Dept. of Agriculture, on the *Poultry and Egg Situation*, March, 1941.)

Year.	Jan.	Feb.	Mar.	Apr.	May.	June.	July.	Aug.	Sept.	Oct.	Nov.	Dec.
1938	7·9	9·9	15·4	17·5	17·3	14·9	13·6	11·8	9·4	7·5	5·9	6·4
1939	8·0	9·7	14·9	17·0	17·0	14·6	13·2	11·7	9·3	7·4	6·0	6·8
1940	7·2	9·0	14·4	16·5	17·0	14·8	13·4	11·8	9·7	7·9	6·2	6·8

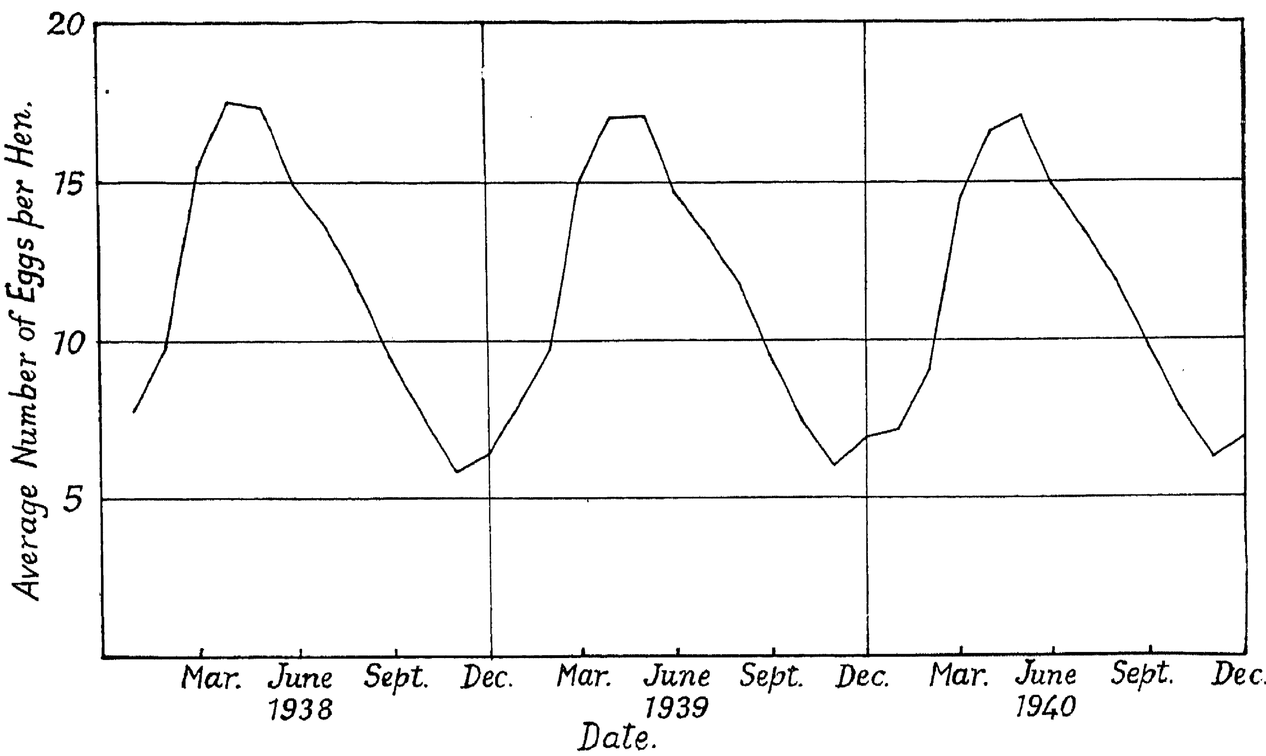


FIG. 29.5.—Graph of the Data of Table 29.5 (Egg Production).

These series are fairly typical of the kind of material with which our theory has to deal. The data of Table 29.1 (barley yields) present a very irregular fluctuation, and so far as the eye can see (which is not a decisive test) there is no systematic oscillation and no regular movement in mean yields over the period. By contrast, Table 29.2 (human population) shows a relatively smooth movement without apparent oscillation. Table 29.3 (sheep population) combines a general decline in numbers with marked oscillatory effects which, though not perfectly regular, appear to be systematic to some extent. Tables 29.4 and 29.5 exhibit an oscillatory effect which is definitely seasonal for the latter and much less regular for the former, neither indicating a variation, in the periods covered, of the average values about which the series oscillate.

29.4. It must not be overlooked that our method of determining the values of the series at fixed equal intervals of time may suppress evidence of oscillatory movements which have a period equal to those intervals or to some sub-multiple of them. Suppose, for instance, that there was a systematic oscillation in the English population expressible

by a harmonic component with period of exactly 10 years, or exactly 5 years, or exactly $3\frac{1}{3}$ years. Clearly, by observing the series at 10-yearly intervals we should never find any evidence of this effect, for it would contribute exactly the same amount to each observation, without oscillation. In the population case, of course, we have collateral evidence to indicate that no such oscillation exists, but where nothing is known of the series otherwise we can never exclude the possibility of a period exactly equivalent to our time-interval. Sometimes, in fact, we know that it is there, and choose our interval so as to exclude the oscillation from consideration. For instance, in our sheep population we know that there is a seasonal effect within the year, which is not brought out in Table 29.2 because the sheep census is taken on June 4th each year; and again, in the rainfall data of Table 29.4 we have taken as representing the year the whole rainfall within the year, knowing quite well that rainfall is seasonal to some extent, even in London.

29.5. A general survey of these and similar series suggests that the typical time-series may be regarded as composed of three parts :—

(*a*) a trend, or long-term movement;
(*b*) an oscillation about the trend of greater or less regularity;
(*c*) a "random", "irregular" or "unsystematic" component.

It is customary to regard the series as composed of these elements superposed one on another; that is to say, we consider the movement of the series as the sum of three different components which may be generated by different causal systems. Particular series, of course, need not exhibit them all. That of Table 29.2 (human population) seems to be almost entirely trend, with perhaps a small unsystematic residual, whereas that of Table 29.5 (egg production) appears to be entirely oscillatory, and very regularly so. But some series at least exhibit all three.

29.6. The primary problem of time-series analysis from the statistical viewpoint is to isolate the three factors for individual study, and in this chapter and the next we shall be mainly concerned with various methods of carrying out the necessary analysis. Before proceeding, however, we must look a little more closely into the reality of the effects which we are investigating and the basis on which we assume that the analysis is legitimate.

29.7. Perhaps the easiest component to understand and to remove from the series is the *seasonal effect*. This is a fluctuation imposed on the series by a cyclic phenomenon external to the main body of causal influences at work upon it. The oscillation in egg-production in Table 29.5, for instance, reflects the rhythm in the reproductive process which is found among birds in virtue, ultimately, of the fact that the earth goes round the sun once a year. Strictly speaking, we ought to confine the word "seasonal" to those effects which are annual in period; but where no confusion is likely to arise we can apply the same word and the same ideas to any phenomenon generated by strictly periodic natural processes, such as "spring" and "neap" variation in tides or daily variation in temperature. We must, however, be careful about extending the notion of seasonality to phenomena which are not demonstrated beyond reasonable doubt to depend on strictly periodic stimuli. For instance, it would be going too far, in the present state of our knowledge, to speak of sunspot variation as seasonal in this sense, and much too far to speak of seasonality in crop-yields as determined by sunspots, even if the relation between the two were established. We shall return to this point below when defining what we mean by a "cycle" as distinct from an "oscillation".

29.8. As we noted in **29.4**, the seasonal effect may already be removed from the series by the way in which the data are specified. Where we ourselves have any choice in the determination of the data, we may eliminate seasonality in the same way, namely, by selecting for measurement of the series a point of time which is fixed in relation to the year, such as June 4th for the agricultural returns of England and Wales, or by averaging over the year, or (what is much the same thing) by cumulating the series over the year, as for instance with rainfall data.

29.9. The concept of *trend* is more difficult to define. Generally, one thinks of it as a smooth broad motion of the system over a long term of years, but " long " in this connection is a relative term, and what is long for one purpose may be short for another. For example, if we were examining rainfall records over a hundred years a slow rise from the beginning of the period to the end would be regarded as a trend ; but if we possessed records for two thousand years (and the rings in some of the giant redwood trees give an index of climatic conditions for periods of this order) the rise over a particular century might appear as part of a slow oscillatory movement, so that any inference from the " trend " in a particular century to the effect that the weather was likely to continue becoming wetter and wetter might be quite false. What inference we should make in practice would depend on what we were trying to do. If we were engineers designing a water-supply system and wished to provide against droughts of reasonable extent, we might perhaps assume that the trend would last as long as our works and proceed accordingly ; but if we were attempting to study climatic changes over the face of the earth for geological periods of time we should accept the continuance of the trend with the greatest reserve or, more probably, should reject it on collateral grounds.

29.10. However long a series may be, we can never be certain, and often not even reasonably sure, that a trend in it is not part of a slow oscillation, except of course when the series has terminated (as might, for instance, be the case if we were considering the lengths of reigns of the Roman Emperors). In speaking of a trend, therefore, we must bear in mind the length of the series to which our statement refers. Perhaps it would be more accurate to speak of slow or quick movements rather than of trend and oscillation, but even so the distinction between the two would remain a matter of subjective judgment to some extent.

29.11. When seasonal variation and trend have been removed from the data we are left with a series which will present, in general, fluctuations of a more or less regular kind. Fig. 29.1 represents the kind of series we obtain, since it has no components of trend or seasonality. The question then arises, is this residual series systematic in the sense that its values can be represented as a function of the time ? Or, on the other hand, are the values random in the sense that they could occur, *in the observed order*, by random sampling from a homogeneous population ? Or again, is there some possibility intermediate between complete functional variation and complete randomness ? The search for systematic effects in residual fluctuation gives rise to several techniques of analysis, the object of which is to detect whether any part of the series is subject to law, and therefore predictable, and whether any part is purely haphazard. The former part we shall call systematic, and it will be referred to as an " oscillation " (not a " cycle ", which is a very special case of an oscillation, as we shall see later). The remainder of the series we shall call the unsystematic component, and refer to its movements as " random ". When a series is a mixture

of oscillation and random movement it will not cause any inconvenience to refer to the up-and-down movement generally as fluctuation before we have analysed it into its constituents ; that is to say, we may speak of fluctuation without prejudice to the possibility of detecting oscillatory movements in it.

In this chapter we study trend and random residuals. In the next chapter we shall deal with oscillatory and cyclical components.

29.12. The logician or the economist who wants to be difficult can always maintain that, although any series can be separated into our three specified components as a matter of mathematical or statistical analysis, the results throw little or no light on the causal influences at work to produce the series. To such a critic we have to concede, I think, that in carrying out the analysis we have at the back of our minds the strong possibility that the three elements are due to independent causal systems. If he refuses to accept this view—and some economists do—we can only invite him to produce a better statistical method.

Possibly the reader will feel, on reaching the end of Chapter 30, that we have not been wasting our time, and that our methods do throw light on the way in which time-series behave. If not, he should consult some of the references and see whether he finds them statistically more satisfying.

Determination of Trend

29.13. It is an essential part of the concept of trend that the movement over fairly long periods is smooth. This means that we can represent the trend component, at least locally, by a polynomial in the time element t. Thus, given the series u_t, we may, in the first instance, seek for some polynomial

$$u_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_p t^p \quad \ldots \quad (29.3)$$

which will give an account of the trend movement. By taking p great enough we can, of course, obtain as close a representation as we like to a finite series ; and how large we take p is a matter for decision in particular cases.

If the polynomial is fitted to the whole series by least squares, it evidently gives the curvilinear regression line of u_t on the variable t. This method would then lead to the fitting of regressions in the manner of Chapter 22, and we need not repeat here what has been said on the subject in that chapter. In Example 22.7 we did, in fact, fit a quartic to the population data of Table 29.2 and found a good fit.

29.14. It is, however, clear that to obtain a satisfactory trend-curve for data such as that of Table 29.3 (sheep population), we should have to take a polynomial of rather high order. This may appear somewhat artificial and in any case the coefficients of such a polynomial, being based on high-order moments, would be very unstable from the sampling viewpoint. A more practical objection, though by no means an unimportant one, is that if we add another term to the series, as for example if we are keeping an annual series up to date from year to year, the work of fitting has to be done afresh each time. Moreover, the trend-line may be affected throughout its length. When, therefore, the series has no very obvious trend such as that of Table 29.2 it is more convenient to use the simpler methods described below.

Moving Averages

29.15. An alternative to finding a polynomial which will represent the whole series is to determine a polynomial which will represent a part of it, and to use different polynomials for different parts. The simplest method, and one which forms the basis of the majority of methods of trend fitting, is to take the first m terms (m being chosen at will), fit a polynomial of order p, not greater than $m - 1$, to them, and use that polynomial to determine the value in the middle of its range; then to repeat the operation with the m terms from the second to the $(m + 1)$th, and so on, moving on one term at each stage. Unless other considerations require it, we take m to be odd, so that the middle point of the range corresponds to a value which is actually observed. Otherwise the middle point falls half-way between two observed values, or we have to use some value of the fitted polynomial other than the middle point, which results in a loss of useful symmetry.

29.16. Suppose, then, that the number of terms is chosen to be odd and is denoted, with a slight change of notation, by $2m + 1$. Without loss of generality we may denote the terms by $u_{-m}, u_{-(m-1)}, \ldots u_0, \ldots u_{m-1}, u_m$. If we choose to fit to them a polynomial of the pth order (29.3) we may, in the usual way, determine the coefficients by least squares, i.e. solve the equations

$$\frac{\partial}{\partial a_j} \sum_{t=-m}^{m} (u_t - a_0 - \ldots - a_p t^p)^2 = 0, \qquad j = 0 \ldots p \quad . \quad . \quad (29.4)$$

which will give us equations typified by

$$\Sigma (t^j u_t) - a_0 \Sigma (t^j) - a_1 \Sigma (t^{j+1}) - \ldots - a_p \Sigma (t^{j+p}) = 0. \quad . \quad . \quad (29.5)$$

Now the sums $\Sigma (t^j)$ are functions of m only. Thus, if we solve (29.5) for a_0 we shall find an equation of the form

$$a_0 = c_0 + c_1 u_{-m} + c_2 u_{-(m-1)} + \ldots + c_{2m+1} u_m, \quad . \quad . \quad (29.6)$$

where the c's depend on m and p, but not on the u's.

Now u_0 assumes the value a_0 at $t = 0$ and hence this value, as given by (29.6), is the value we require for the polynomial. As we see, this is equivalent to a weighted average of the observed values, the weights being independent of which part of the series is taken. Thus our process of fitting a trend-line consists of determining the constants c (which depend on m and p and therefore give us a twofold element of choice) and then calculating, for each consecutive set of $(2m + 1)$ terms in the series, a value given by (29.6). If the terms are $u_x \ldots u_{2m+x}$, the calculated value will correspond to $t = m + x$. There will be no values corresponding to the m terms at the beginning and the m terms at the end.

Example 29.1

Suppose we have a series and wish to fit a curve which best approximates to sets of seven points; and suppose we regard a cubic as providing a satisfactory approximation. What are the weights of the moving average?

We have $m = 3$ and $p = 3$, and our polynomial is

$$u_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

Taking our origin at $t = 0$, we find, for equations (29.5), in virtue of the fact that $\Sigma(t^k) = 0$ for odd k,

$$\begin{aligned}
\Sigma(u) &= 7a_0 \qquad\qquad + 28a_2 \\
\Sigma(tu) &= \qquad 28a_1 \qquad\qquad + 196a_3 \\
\Sigma(t^2u) &= 28a_0 \qquad\quad + 196a_2 \\
\Sigma(t^3u) &= \qquad 196a_1 \qquad\qquad + 1588a_3
\end{aligned}$$

giving, for a_0,

$$a_0 = \frac{1}{21}\{7\Sigma(u) - \Sigma(t^2u)\}$$

$$= \frac{1}{21}\{-2u_{-3} + 3u_{-2} + 6u_{-1} + 7u_0 + 6u_1 + 3u_2 - 2u_3\}.$$

We may write this conveniently as

$$\frac{1}{21}[-2,\ 3,\ 6,\ 7,\ 6,\ 3,\ -2]$$

or, when symmetrical formulae are used, as in the present case, by

$$[-2,\ 3,\ 6,\ \mathbf{7}\ .\ .\ .\ ],$$

denoting the middle term by heavy type.

To take a simple illustration. Suppose the series is given by the following values :—

t :	1	2	3	4	5	6	7	8	9	10
u_t :	0	1	8	27	64	125	216	343	512	729

We have, for the trend value at $t = 4$,

$$a_0 = \frac{1}{21}\{(-2\times 0)+(3\times 1)+(6\times 8)+(7\times 27)+(6\times 64)+(3\times 125)-(2\times 216)\} = \frac{1}{21}\{567\}$$

$$= 27.$$

Similarly, at $t = 6$ we find

$$a_0 = \frac{1}{21}\{(-2\times 8) + (3\times 27) + \ .\ .\ .\ -(2\times 512)\}$$

$$= 125.$$

In both cases the trend-value is equal to the actual value of the series, and this obviously must be so when we note that we are fitting a cubic to the series

$$u_t = (t-1)^3.$$

It will be observed that in this example we should have obtained the same value for a_0 if we fitted quadratics instead of cubics; and generally the case p odd includes the case of the next lowest (even) value of p, so that we need not give separate formulae for even p.

29.17. Writing $a_0[k]$ for the value of a_0 calculated in the above manner for an average of k successive terms, we find the following formulae up to $p = 5$. The reader may care to verify them for himself as an exercise.

Quadratic and Cubic

$$\left.\begin{array}{lll}
a_0 & [5] & \frac{1}{35}[-3,\ 12,\ \mathbf{17},\ \ldots] \\
 & [7] & \frac{1}{21}[-2,\ 3,\ 6,\ \mathbf{7},\ \ldots] \\
 & [9] & \frac{1}{231}[-21,\ 14,\ 39,\ 54,\ \mathbf{59},\ \ldots] \\
 & [11] & \frac{1}{429}[-36,\ 9,\ 44,\ 69,\ 84,\ \mathbf{89},\ \ldots] \\
 & [13] & \frac{1}{143}[-11,\ 0,\ 9,\ 16,\ 21,\ 24,\ \mathbf{25},\ \ldots] \\
 & [15] & \frac{1}{1105}[-78,\ -13,\ 42,\ 87,\ 122,\ 147,\ 162,\ \mathbf{167},\ \ldots] \\
 & [17] & \frac{1}{323}[-21,\ -6,\ 7,\ 18,\ 27,\ 34,\ 39,\ 42,\ \mathbf{43},\ \ldots] \\
 & [19] & \frac{1}{2261}[-136,\ -51,\ 24,\ 89,\ 144,\ 189,\ 224,\ 249,\ 264,\ \mathbf{269},\ \ldots] \\
 & [21] & \frac{1}{3059}[-171,\ -76,\ 9,\ 84,\ 149,\ 204,\ 249,\ 284,\ 309,\ 324,\ \mathbf{329},\ \ldots]
\end{array}\right\} \tag{29.7}$$

Quartic and Quintic

$$\left.\begin{array}{ll}
[7] & \frac{1}{231}[5,\ -30,\ 75,\ \mathbf{131},\ \ldots] \\
[9] & \frac{1}{429}[15,\ -55,\ 30,\ 135,\ \mathbf{179},\ \ldots] \\
[11] & \frac{1}{429}[18,\ -45,\ -10,\ 60,\ 120,\ \mathbf{143},\ \ldots] \\
[13] & \frac{1}{2431}[110,\ -198,\ -135,\ 110,\ 390,\ 600,\ \mathbf{677},\ \ldots] \\
[15] & \frac{1}{46{,}189}[2145,\ -2860,\ -2937,\ -165,\ 3755,\ 7500,\ 10{,}125,\ \mathbf{11{,}063},\ \ldots] \\
[17] & \frac{1}{4199}[195,\ -195,\ -260,\ -117,\ 135,\ 415,\ 660,\ 825,\ \mathbf{883},\ \ldots ,] \\
[19] & \frac{1}{7429}[340,\ -255,\ -420,\ -290,\ 18{,}405,\ 790,\ 1110,\ 1320,\ \mathbf{1393},\ \ldots] \\
[21] & \frac{1}{260{,}015}[11{,}628,\ -6460,\ -13{,}005,\ -11{,}220,\ -3940,\ 6378,\ 17{,}655, \\
 & \qquad 28{,}190,\ 36{,}660,\ 42{,}120,\ \mathbf{44{,}003},\ \ldots]
\end{array}\right\} \tag{29.8}$$

29.18. Several methods have been proposed to simplify the arithmetic of fitting a trend-line by moving averages, the large numbers in some of the expressions in (29.7) and (29.8) involving considerable labour in straightforward application. The simplest, perhaps, is that of iterated averages.

Suppose we take an average of sets of four with equal weights—a very simple process

—and then another average of the same kind *of that average.* If the primary series is u_t, the result of the first operation will be to give a series

$$v_1 = \frac{1}{4}(u_1 + u_2 + u_3 + u_4)$$

$$v_2 = \frac{1}{4}(u_2 + u_3 + u_4 + u_5), \text{ etc.,}$$

and that of the second operation to give

$$\begin{aligned} w_1 &= \frac{1}{4}(v_1 + v_2 + v_3 + v_4) \\ &= \frac{1}{16}[u_1 + 2u_2 + 3u_3 + 4u_4 + 3u_5 + 2u_6 + u_7]. \qquad (29.9) \end{aligned}$$

We may write this symbolically as

$$\left\{\frac{1}{4}[1, 1, 1, 1]\right\}^2 = \frac{1}{16}[1, 2, 3, \mathbf{4} \ldots], \qquad (29.10)$$

or, reserving the symbol $\frac{1}{k}[k]$ for a simple arithmetic mean of k terms, as

$$\frac{1}{16}[4]^2 = \frac{1}{16}[1, 2, 3, \mathbf{4} \ldots]. \qquad (29.11)$$

Now compare the weights of the average derived in Example 29.1 for fitting a cubic to seven points. Reduced to unit divisors we have for the weights of the latter

$$-0{\cdot}0952,\ 0{\cdot}1429,\ 0{\cdot}2857,\ \mathbf{0{\cdot}3333} \ldots$$

and for the weights of (29.9)

$$0{\cdot}0625,\ 0{\cdot}1250,\ 0{\cdot}1875,\ \mathbf{0{\cdot}2500} \ldots$$

The two are not identical, but they follow the same sort of course and it might be possible to regard the latter as an approximation to the former. (We shall derive better approximations presently, but this will serve for purposes of illustration.) Now the iterated summation resulting in (29.9) is much easier to carry out than the single weighted averaging process of Example 29.1. Generally, if we can find averages with simple integral weights, preferably unity, which will, in conjunction, give approximations to the more complicated weights of a single average, it is usually easier to use the iteration process.

29.19. In the notation of finite differences, write

$$\Delta u_t = u_{t+1} - u_t \qquad (29.12)$$

$$E\,u_t = u_{t+1} = (1 + \Delta)\,u_t \qquad (29.13)$$

$$\delta u_t = u_{t+\frac{1}{2}} - u_{t-\frac{1}{2}}. \qquad (29.14)$$

We have, for the second "central" difference $\delta^2 u_t$,

$$\begin{aligned} \delta^2 u_t &= (u_{t+1} - u_t) - (u_t - u_{t-1}) \\ &= (E - 2 + E^{-1})\,u_t. \qquad (29.15) \end{aligned}$$

Writing

$$E = \exp(2i\phi) \qquad (29.16)$$

we find, symbolically,

$$\begin{aligned} \delta^2 &= E - 2 + E^{-1} \\ &= \exp(2i\phi) + \exp(-2i\phi) - 2 \\ &= -4\sin^2\phi. \qquad (29.17) \end{aligned}$$

Then
$$\sum_{t=-m}^{m}(u_t) = \sum_{t=-m}^{m}(E^t u_0)$$
$$= \left\{1 + 2\sum_{j=1}^{m}(\cos 2j\phi)\right\} u_0,$$
since the terms in $\sin 2j\phi$ vanish,
$$= \frac{\sin(2m+1)\phi}{\sin\phi} u_0. \qquad . \quad . \quad . \quad . \quad . \quad (29.18)$$
Thus
$$\frac{1}{k}[k]\,u_0 = \frac{1}{k}\frac{\sin k\phi}{\sin\phi} u_0$$
$$= \frac{1}{k}\left\{k - \frac{k(k^2-1)}{3\,!}\sin^2\phi + \frac{k(k^2-1^2)(k^2-3^2)}{5\,!}\sin^4\phi - \dots\right\} u_0$$
$$= u_0 + \frac{k^2-1}{2^2\,3\,!}\delta^2 u_0 + \frac{(k^2-1)(k^2-3^2)}{2^4\,5\,!}\delta^4 u_0 + \dots \qquad . \quad . \quad (29.19)$$
This interesting formula gives the arithmetic average in terms of the middle term u_0 and its central differences.

If now our series is approximately represented by a cubic, so that fourth differences vanish, we have
$$\frac{1}{k}[k]\,u_0 = u_0 + \frac{k^2-1}{24}\delta^2 u_0 \qquad . \quad . \quad . \quad . \quad (29.20)$$
and this equation will in any case be true up to third differences. Similarly, for two iterated averages we have, to the same order,
$$\frac{1}{k_1 k_2}[k_1][k_2]\,u_0 = u_0 + \frac{1}{24}(k_1^2 + k_2^2 - 2)\,\delta^2 u_0 \qquad . \quad . \quad . \quad (29.21)$$
and so on. We will use these results to derive two formulae in very general use by actuaries for " graduating " a series, a process which is very similar to that of fitting a trend-line.

Example 29.2. Spencer's 15-point Formula

Consider three successive averages with equal weights
$$\frac{1}{80}[4][4][5]\,u_0 = u_0 + \frac{1}{24}\{4^2 - 1 + 4^2 - 1 + 5^2 - 1\}\,\delta^2 u_0$$
$$= u_0 + \frac{9}{4}\delta^2 u_0.$$
We then have, to third differences
$$u_0 = \frac{1}{80}[4]^2[5]\left(1 - \frac{9}{4}\delta^2\right) u_0.$$
Substituting for δ^2 the formula [1, − 2, 1], as given by (29.15), we find
$$u_0 = \frac{1}{320}[4]^2[5][-9,\ 22,\ -9].$$
Now without affecting the order of the approximation we may add factors in δ^4 or higher central differences, and can simplify the numerical coefficients to some extent. Let us

add to the factor $[-9, 22, -9]$ a term $-3\delta^4 = [-3, 12, -18, 12, -3]$. The result is $[-3, 3, 4, 3, -3]$, giving

$$u_0 = \frac{1}{320}[4]^2[5][-3, 3, \mathbf{4}, \ldots].$$

This is Spencer's 15-point formula. It covers sets of 15 consecutive terms, the weights in full being

$$\frac{1}{320}[-3, -6, -5, 3, 21, 46, 67, \mathbf{74}, \ldots]$$

Example 29.3. Spencer's 21-point Formula

In a similar way we find

$$\frac{1}{175}[5]^2[7] = 1 + 4\delta^2,$$

giving, to third differences,

$$u_0 = \frac{1}{175}[5]^2[7](1 - 4\delta^2)u_0$$

$$= \frac{1}{175}[5]^2[7][-4, 9, -4]u_0.$$

We now add to the factor $[-4, 9, -4]$ the expression

$$-3\delta^4 - \tfrac{1}{2}\delta^6 = [-3, 12, -18, 12, -3] + [-\tfrac{1}{2}, 3, -7\tfrac{1}{2}, 10, -7\tfrac{1}{2}, 3, -\tfrac{1}{2}]$$

giving

$$u_0 = \frac{1}{175}[5]^2[7][-\tfrac{1}{2}, 0, \tfrac{1}{2}, 1, \tfrac{1}{2}, 0, -\tfrac{1}{2}]$$

$$= \frac{1}{350}[5]^2[7][-1, 0, 1, \mathbf{2}, \ldots].$$

This is Spencer's 21-point formula.

29.20. A few practical points arising in the application of the foregoing formulae are worth mentioning.

(*a*) The order in which the iterations are carried out is of course immaterial, as the reader can easily verify. It is therefore more convenient, as a rule, to carry out the more complicated operations first, while the numbers being handled remain small. For instance, in applying the Spencer 15-point formula we should carry out the moving average $[-3, 3, 4, 3, -3]$ first, then apply the simple average $\frac{1}{5}[5]$, and then the two averages of four. This does not apply if the series is short, inasmuch as there are fewer of the final than of the initial operations.

(*b*) The use of a moving average of extent $2k + 1$ involves the absence of k terms at the end and k terms at the beginning of the trend-series. If the original series is short the loss may be serious, and this effect sometimes restricts considerably the extent of the average which we are able to apply.

(*c*) It is possible to remedy the deficiency at the ends of the series by special formulae, but the values so derived have less reliability than those of the main trend-line, and on the whole it seems better to accept the loss of $2k$ terms unless trend-values for the beginning and end of the series are really essential.

(*d*) As yet we have given no guide as to the choice of most suitable values of m and p. In practice we do not usually require to fit curves of degree higher than five, and often a cubic is sufficient, as is assumed in the Spencer formulae. There is greater elasticity in the choice of m, but the point mentioned in (*b*) above requires m to be as small as possible, consistent with other requirements. We shall see later in the chapter that the variate-difference method gives some further guide as to p, and that certain effects of trend-elimination on random elements bear on the extent determined by m.

(*e*) There is a voluminous literature on trend-fitting which appears to me out of proportion to the importance of the subject. It is not difficult to pursue inquiries on the above lines to the point of extreme apparent precision and great mathematical complexity, and perhaps such work is valuable where the series is fairly smooth and not disturbed seriously by sampling variation or superposed random fluctuation. But many of the series encountered in statistical practice will not bear the weight of great refinement in trend-fitting. The student will probably find that a knowledge of fitting by moving averages will be sufficient for all ordinary and many extra-ordinary purposes.

The Effect of Trend-elimination on Other Components

29.21. In Table 29.6 we have applied the Spencer 21-point formula to an artificial series obtained by adding a random element to a cubic. Specifically,

$$u_t = (t - 26) + \frac{1}{10}(t - 26)^2 + \frac{1}{100}(t - 26)^3 + \varepsilon_t. \qquad (29.22)$$

The component ε_t was taken from tables of random numbers and consists of samples from a population in which all integral values from 0 to 99 are equally frequent. The various columns of the table illustrate the process of fitting, and we may note in passing that for a series as short as this it is convenient to leave the more difficult summations to the last as there are substantially fewer of them.

Now we know that the Spencer formula will fit a cubic exactly, so that when we subtract the trend from the original series we ought to eliminate the systematic constituent entirely and be left with our random component, except in so far as we have rounded off the systematic element to the nearest unit. A comparison of columns (2) and (9) in Table 29.6, remembering that the latter includes an element 49·5 equal to the mean of the random component, shows that we do not do so. The reason is not far to seek. The moving average has acted on the random element itself and determined a trend-line in it.

The results of applying the Spencer 21-point formula to the random element ε_t are shown in column (11). We should expect that if the method were perfect the values in this column would be 49·5, the mean of ε_t, apart from irregular sampling effects; but not only do the observed values deviate from this mean, they do so systematically, the values having a small oscillatory movement which is shown as part of the trend.

29.22. This effect can assume considerable importance, particularly if we are eliminating trend so as to concentrate attention on oscillations. We proceed to examine it more closely.

Suppose that we have a series composed of the sum of three parts, a trend $\phi_1(t)$, an oscillatory term $\phi_2(t)$, and a random element $\phi_3(t)$, so that

$$u_t = \phi_1 + \phi_2 + \phi_3. \qquad (29.23)$$

TABLE 29.6

Series given by Equation (29.22) with Trend-Line determined by a Spencer 21-point Formula.

(1) t	(2) Cubic Term.	(3) ε_t	(4) u_t	(5) $[5]\,u_t$	(6) $[5]\,(5)$.	(7) $[7]\,(6)$.	(8) $[-1, 0, 1, 2\;.\;.\;]\,(7)$.	(9) $\frac{1}{350}$ (8).	(10) Deviation $u_t-(9)$.	(11) Graduation of ε_t alone.
1	−119	23	−96	...	...	...	...	...	...	...
2	−105	15	−90	...	...	...	...	...	...	...
3	− 92	75	−17	−246	...	...	...	...	...	...
4	− 80	48	−32	−209	...	...	...	...	...	...
5	− 70	59	−11	− 87	−572	...	...	...	...	...
6	− 60	1	−59	− 42	−241	...	...	...	...	...
7	− 51	83	32	12	162	...	...	...	...	...
8	− 44	72	28	85	413	2,233	...	...	...	...
9	− 37	59	22	194	670	3,801	...	...	...	...
10	− 31	93	62	164	844	5,120	...	...	...	...
11	− 26	76	50	215	957	5,984	14,352	41	9	67
12	− 22	24	2	186	996	6,642	15,470	44	−42	66
13	− 18	97	79	198	1,078	7,041	15,815	45	34	63
14	− 15	8	− 7	233	1,026	7,145	15,676	45	−52	60
15	− 12	86	74	246	1,071	7,038	14,978	43	31	55
16	− 10	95	85	163	1,069	6,934	14,166	40	45	51
17	− 8	23	15	231	948	6,709	13,379	38	−23	47
18	− 7	3	− 4	196	850	6,535	12,703	36	−40	43
19	− 6	67	61	112	892	6,408	12,169	35	26	40
20	− 5	44	39	148	853	6,363	12,102	35	4	39
21	− 4	5	1	205	852	6,446	12,279	35	−34	39
22	− 3	54	51	192	944	6,611	12,676	36	15	39
23	− 2	55	53	195	1,024	6,769	13,228	38	15	40
24	− 2	50	48	204	1,031	7,052	13,857	40	8	41
25	− 1	43	42	228	1,015	7,353	14,508	41	1	42
26	0	10	10	212	1,050	7,610	15,120	43	−33	43
27	1	74	75	176	1,136	7,923	15,634	45	30	44
28	2	35	37	230	1,153	8,249	16,251	46	− 9	44
29	4	8	12	290	1,201	8,607	17,002	49	−37	45
30	6	90	96	245	1,337	9,019	17,717	51	45	44
31	9	61	70	260	1,357	9,424	18,499	53	17	44
32	12	18	30	312	1,373	9,870	19,307	55	−25	43
33	15	37	52	250	1,462	10,429	20,159	58	− 6	42
34	20	44	64	306	1,541	10,989	21,133	60	4	41
35	24	10	34	334	1,599	11,679	22,417	64	−30	39
36	30	96	126	339	1,760	12,539	23,797	68	58	38
37	36	22	58	370	1,897	13,529	25,737	74	−16	37
38	44	13	57	411	2,047	14,699	27,955	80	−23	36
39	52	43	95	443	2,233	16,060	30,456	87	8	35
40	61	14	75	484	2,452	17,570	33,334	95	−20	34
41	71	87	158	525	2,711	19,353	36,716	105	53	34
42	83	16	99	589	2,960	21,394	...	...	...	...
43	95	3	98	670	3,270	23,690	...	...	...	...
44	109	50	159	692	3,680	26,255	...	...	...	...
45	124	32	156	794	4,088	...	...	...	...	...
46	140	40	180	935	4,529	...	...	...	...	...
47	158	43	201	997	5,017	...	...	...	...	...
48	177	62	239	1,111	...	...	...	...	...	...
49	198	23	221	1,180	...	...	...	...	...	...
50	220	50	270	...	...	...	...	...	...	...
51	244	5	249	...	...	...	...	...	...	...

If we determine the trend by a moving average, denoted by an operation T, then clearly

$$Tu_t = T\phi_1 + T\phi_2 + T\phi_3. \qquad (29.24)$$

Let us now suppose that our method of determining trend is perfect in the sense that $T\phi_1 = \phi_1$. Then, on subtracting (29.24) from (29.23) to eliminate trend, we find

$$u_t - Tu_t = (\phi_2 - T\phi_2) + (\phi_3 - T\phi_3). \qquad (29.25)$$

The point of present interest is that the terms $T\phi_2$ and $T\phi_3$ in (29.25) may distort the genuinely oscillatory parts of the residual series and induce spurious oscillatory movements.

29.23. Consider the simple case when ϕ_2 is a sine term, $\sin(\alpha + \lambda t)$, t being integral. Since

$$\sum_{t=1}^{k} \sin(\alpha + \lambda t) = \frac{\sin \frac{1}{2}k\lambda}{\sin \frac{1}{2}\lambda} \sin \{\alpha + \tfrac{1}{2}(k+1)\lambda\}, \qquad (29.26)$$

a simple moving average of k consecutive terms will result in a sine series of the same period and phase as the original, but with the amplitude reduced by the factor

$$\frac{1}{k}\frac{\sin \frac{1}{2}k\lambda}{\sin \frac{1}{2}\lambda}. \qquad (29.27)$$

Iteration q times will reduce the amplitude by the qth power of this factor.

Thus the term $T\phi_2$ will be small if k is large, q is large, or if $\frac{1}{2}k\lambda$ is a multiple of π, that is, if the extent of the moving average is a period of the oscillation. But if λ is small and $k\lambda$ is small the amplitude is reduced very little and $\phi_2 - T\phi_2$ will largely disappear, i.e. the moving average will partially obliterate the term in ϕ_2. In this case, $k\lambda$ being small, the extent of the moving average is small compared with the period of the harmonic term, that is to say the oscillation is a slow one. This result is what we should expect. A slow oscillation is treated as a trend by the moving average and eliminated accordingly. Generally, the moving average will emphasise the shorter oscillations at the expense of the longer ones. Furthermore, if the extent of the average is slightly greater than the period, the term (29.27) may have a negative sign, and consequently the difference from the trend may somewhat exaggerate the true oscillations.

It is not so easy to exhibit the precise effect of the moving average when the weights are unequal and the terms are not harmonic, but evidently the same kind of situation is apt to arise.

29.24. Now consider the effect of a simple moving average (that is, one with equal weights) on the residual element ϕ_3 which we will suppose to be a random element ε_t with variance v. For the term $T\phi_3$ we have

$$T\phi_3 = \frac{1}{k}\sum_{-[\frac{1}{2}k]}^{[\frac{1}{2}k]} \varepsilon_{t+j} \qquad (29.28)$$

where $[\frac{1}{2}k]$ is the greatest integer which does not exceed $\frac{1}{2}k$. Consecutive values of ε_t are independent, but consecutive values of $T\phi_3$ are not; for $T\phi_3(a)$ and $T\phi_3(b)$ have $k - (a - b)$ values of ε in common and are correlated if $a - b < k$. Thus the series $T\phi_3$ will be much smoother than ϕ_3, and if we proceed to further averagings will become smoother still. We have had an example of this effect in Table 29.6, and shall meet further examples below.

29.25. The effect of taking a moving average of a random series will then be to generate an oscillatory series, provided that the weights are such as to give a positive correlation between successive members of the generated series, a condition which is always realised in moving averages employed for trend-fitting. We shall call this the Slutzky–Yule effect, after the two statisticians who (independently) studied it in detail.

The generated series is not regular in the cyclical sense, that is to say its peaks and troughs do not recur at equal intervals of time, and the amplitudes of the oscillations vary considerably. Nevertheless such oscillations present a striking resemblance to the kind of movement which is found in practice, particularly in economic time-series, and we shall consider them in more detail in Chapter 30. For our present purposes we require to consider how far the process of trend-elimination itself may generate such effects in order to be sure that oscillatory movements in a trend-free series have not been put there, so to speak, by our own arithmetical processes.

29.26. For this purpose we shall consider the period and variance of a series generated by the Slutzky–Yule effect.

Since the peaks and troughs do not recur at equal intervals there is no quantity which we can conveniently call the length of the oscillation. There will, in fact, be a distribution of lengths. We may define as the mean length either the mean period from peak to peak, or that from trough to trough; but this raises some difficulties as to whether we are prepared to admit as periods small ripples on the main undulation.

Recognising its somewhat arbitrary character, we shall take as our measure of oscillatory length the mean distance between "upcrosses", that is to say the mean distance between points where the series changes sign from negative to positive or "crosses the x-axis". Suppose the series is generated by a moving average with weights $a_1 \ldots a_k$ of a random variable which is normally distributed with variance v. Then the probability that

$$u_k = \sum_{j=1}^{k} a_j \varepsilon_j < 0 \quad . \quad . \quad . \quad . \quad . \quad (29.29)$$

and

$$u_{k+1} = \sum_{j=1}^{k} a_j \varepsilon_{j+1} > 0, \quad . \quad . \quad . \quad . \quad . \quad (29.30)$$

i.e. that the generated series changes sign from negative to positive, is the proportional frequency of

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}(k+1)}} \exp\left[-\frac{1}{2v}\sum_{j=1}^{k+1} \varepsilon_j^2\right] d\varepsilon_1 \ldots d\varepsilon_{k+1} \quad . \quad . \quad (29.31)$$

between the hyperplanes $\sum_{j=1}^{k} a_j \varepsilon_j = 0$ and $\sum_{j=1}^{k} a_j \varepsilon_{j+1} = 0$. This is equal to the angle between these two planes, which is given by

$$\cos\theta = \frac{\sum_{j=1}^{k-1} a_j a_{j+1}}{\sum_{j=1}^{k} a_j^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29.32)$$

Hence the mean distance between upcrosses is $2\pi/\theta$, where θ is given by (29.32).

29.27. In a similar way, the probability that

$$u_{k+1} - u_k < 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (29.33)$$

$$u_k - u_{k-1} > 0, \quad . \quad . \quad . \quad . \quad . \quad (29.34)$$

that is that u_k is a peak of the series, is the angle between the two hyperplanes

$$\sum_{j=1}^{k} a_j\,\varepsilon_{j+1} - \sum_{j=1}^{k} a_j\,\varepsilon_j = 0 \quad . \quad . \quad . \quad . \quad . \quad (29.35)$$

$$\sum_{j=1}^{k} a_j\,\varepsilon_j - \sum_{j=1}^{k} a_j\,\varepsilon_{j-1} = 0 \quad . \quad . \quad . \quad . \quad . \quad (29.36)$$

and is given by

$$\cos\theta_1 = \frac{\begin{array}{l}(a_2 - a_1)\,a_1 + (a_3 - a_2)\,(a_2 - a_1) + \ldots \\ \quad + (a_k - a_{k-1})\,(a_{k-1} - a_{k-2}) - a_k\,(a_k - a_{k-1})\end{array}}{\{a_1^2 + (a_2 - a_1)^2 + \ldots + a_k^2\}}. \quad . \quad (29.37)$$

Thus the mean distance between peaks is $2\pi/\theta_1$. The same formula obviously applies to mean distance between troughs.

29.28. If we wish to exclude "ripples" of a certain length d from consideration we may inquire for the probability that (29.35) and (29.36) are satisfied in conjunction with

$$u_k > u_{k+d}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (29.38)$$

This is evidently the area cut off on the unit sphere by the three planes (29.35), (29.36) and

$$\sum_{j=1}^{k} a_j\,\varepsilon_j - \sum_{j=1}^{k} a_j\,\varepsilon_{j+d} = 0. \quad . \quad . \quad . \quad . \quad . \quad (29.39)$$

If the angles between the planes are A, B and C this area is $A + B + C - 2\pi = \theta_2$, say. The mean length between peaks, ripples excepted, is then $4\pi/\theta_2$.

Example 29.4

In Table 29.7 we show 480 terms of a series of random numbers which can take integral values from 0 to 19, together with a moving sum of fives of a moving sum of threes. Fig. 29.6 shows a portion of the derived series graphically. There are 474 terms of the smoothed series.

The mean value of our series is $15 \times 9{\cdot}5 = 142{\cdot}5$. The number of upcrosses will be found from the table to be 23, the first between the 19th and 20th term of the smoothed series, the last between the 459th and the 460th. The mean distance between upcrosses is then $440/22 = 20$ units. How does this compare with the mean-distance given by "normal" theory?

The weights of the graduation are [1, 2, 3, 3, 3, 2, 1] and from (29.32) we have

$$\cos\theta = \frac{(1 \times 2) + (2 \times 3) + \ldots + (2 \times 1)}{1^2 + 2^2 + \ldots + 1^2}$$

$$= \frac{34}{37} = 0{\cdot}9189$$

$$\theta = 23^\circ\ 14'.$$

Hence the mean distance $= \dfrac{360}{23{\cdot}233} = 15{\cdot}5$ units.

TABLE 29.7

Series of 480 Terms of a Rectangular Random Series ε and a [5] [3] smoothing S.

t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S	t	ε	S
1	3		49	2	61	97	8	112	145	18	136	193	16	147	241	1	99	289	16	188	337	13	107	385	18	143	433	14	172
2	15		50	3	71	98	1	108	146	18	140	194	6	144	242	17	80	290	2	180	338	6	134	386	5	166	434	14	155
3	15		51	10	84	99	5	123	147	6	131	195	10	132	243	0	75	291	10	158	339	15	151	387	17	170	435	9	133
4	8	164	52	5	91	100	13	131	148	0	121	196	5	128	244	3	73	292	12	147	340	7	161	388	12	179	436	8	107
5	19	147	53	10	92	101	11	150	149	4	120	197	7	122	245	0	94	293	15	145	341	13	160	389	7	179	437	3	75
6	1	143	54	3	101	102	14	151	150	11	137	198	8	126	246	6	124	294	5	148	342	13	162	390	14	184	438	1	53
7	3	145	55	2	119	103	6	140	151	15	162	199	18	120	247	17	169	295	6	145	343	5	155	391	15	190	439	3	55
8	12	165	56	11	141	104	13	120	152	15	179	200	0	121	248	16	195	296	15	134	344	15	153	392	11	194	440	1	72
9	19	175	57	14	166	105	1	119	153	11	188	201	7	105	249	17	204	297	6	137	345	10	162	393	13	201	441	5	91
10	13	196	58	18	190	106	4	120	154	9	184	202	9	99	250	15	191	298	13	136	346	3	174	394	18	199	442	16	96
11	16	191	59	8	212	107	13	133	155	12	183	203	5	93	251	3	175	299	2	136	347	18	176	395	6	193	443	8	91
12	4	178	60	14	211	108	13	147	156	18	175	204	3	95	252	14	150	300	14	129	348	19	177	396	19	178	444	2	78
13	17	159	61	15	204	109	8	172	157	7	174	205	12	91	253	9	144	301	9	128	349	8	174	397	13	173	445	0	75
14	8	150	62	17	191	110	12	186	158	15	160	206	4	93	254	11	131	302	4	115	350	5	173	398	1	178	446	2	85
15	6	134	63	7	185	111	12	195	159	5	157	207	2	97	255	3	135	303	14	100	351	16	159	399	13	178	447	7	109
16	15	118	64	9	166	112	19	204	160	11	141	208	11	97	256	15	125	304	2	93	352	7	157	400	18	183	448	17	124
17	3	101	65	11	160	113	13	203	161	9	140	209	6	107	257	1	138	305	0	96	353	16	157	401	17	191	449	12	124
18	3	88	66	14	167	114	11	184	162	14	122	210	7	115	258	14	142	306	8	109	354	8	169	402	3	205	450	5	117
19	7	87	67	5	188	115	18	156	163	1	117	211	6	128	259	9	162	307	12	131	355	6	168	403	18	198	451	2	106
20	4	100	68	17	203	116	2	135	164	11	94	212	15	125	260	13	166	308	10	159	356	15	165	404	14	192	452	2	97
21	5	126	69	18	204	117	4	121	165	1	98	213	4	130	261	11	182	309	11	178	357	19	153	405	14	191	453	15	92
22	14	140	70	13	205	118	10	111	166	8	93	214	13	126	262	15	190	310	17	187	358	4	150	406	13	197	454	8	100
23	15	147	71	18	185	119	8	116	167	2	106	215	4	125	263	8	203	311	10	200	359	5	133	407	5	205	455	2	111
24	10	150	72	0	171	120	10	131	168	18	103	216	7	123	264	17	210	312	12	212	360	9	120	408	19	204	456	4	120
25	3	153	73	19	149	121	3	145	169	1	121	217	13	119	265	19	214	313	17	216	361	12	117	409	17	202	457	11	121
26	10	156	74	1	146	122	16	156	170	7	117	218	4	111	266	10	211	314	17	211	362	2	127	410	18	192	458	15	119
27	13	165	75	14	130	123	12	173	171	9	127	219	13	101	267	17	188	315	16	192	363	11	118	411	5	174	459	8	110
28	14	175	76	12	135	124	8	175	172	13	120	220	0	91	268	11	163	316	9	173	364	12	112	412	7	140	460	3	98
29	15	168	77	2	139	125	19	160	173	2	137	221	3	82	269	9	146	317	9	151	365	5	105	413	15	107	461	1	98
30	8	160	78	7	160	126	11	145	174	16	139	222	11	76	270	1	153	318	2	152	366	0	100	414	1	86	462	4	121
31	10	154	79	16	175	127	1	129	175	1	145	223	0	75	271	11	154	319	16	160	367	12	84	415	4	66	463	13	150
32	1	156	80	15	188	128	4	123	176	17	142	224	10	72	272	17	162	320	10	185	368	6	88	416	2	58	464	17	170
33	18	154	81	17	197	129	16	108	177	13	145	225	1	86	273	17	155	321	17	196	369	2	96	417	3	50	465	19	176
34	17	165	82	11	200	130	3	115	178	0	149	226	4	92	274	4	154	322	15	209	370	4	104	418	2	62	466	5	169
35	4	164	83	6	206	131	13	108	179	15	157	227	6	109	275	8	137	323	13	194	371	15	109	419	10	78	467	4	149
36	10	159	84	17	215	132	0	118	180	7	166	228	18	116	276	2	134	324	10	179	372	6	130	420	0	105	468	15	136
37	16	138	85	18	228	133	10	112	181	16	167	229	3	139	277	18	141	325	18	151	373	5	148	421	9	126	469	8	137
38	2	137	86	19	230	134	4	122	182	16	171	230	7	149	278	8	172	326	0	133	374	14	156	422	16	146	470	6	136
39	13	131	87	15	220	135	19	113	183	7	169	231	12	149	279	9	184	327	9	112	375	14	164	423	9	152	471	14	133
40	3	140	88	13	198	136	3	110	184	6	174	232	15	141	280	19	185	328	8	108	376	11	180	424	11	144	472	9	126
41	14	135	89	8	175	137	4	100	185	13	168	233	11	137	281	17	167	329	3	105	377	8	187	425	12	124	473	0	125
42	7	146	90	10	159	138	7	103	186	17	170	234	1	134	282	4	150	330	9	111	378	15	174	426	2	106	474	15	109
43	16	141	91	14	158	139	0	106	187	14	170	235	8	128	283	8	123	331	12	107	379	18	151	427	3	106	475	7	103
44	3	139	92	5	158	140	16	107	188	2	159	236	9	130	284	5	115	332	3	101	380	7	127	428	9	119	476	5	96
45	10	117	93	12	159	141	13	102	189	15	140	237	14	132	285	6	131	333	8	85	381	1	99	429	6	139	477	1	95
46	12	96	94	18	153	142	0	103	190	9	139	238	9	128	286	7	168	334	5	77	382	2	88	430	17	159	478	11	
47	0	75	95	1	145	143	2	114	191	1	145	239	6	122	287	19	186	335	1	75	383	7	89	431	15	174	479	5	
48	3	65	96	14	124	144	4	127	192	15	151	240	7	108	288	19	196	336	2	93	384	4	119	432	5	179	480	5	

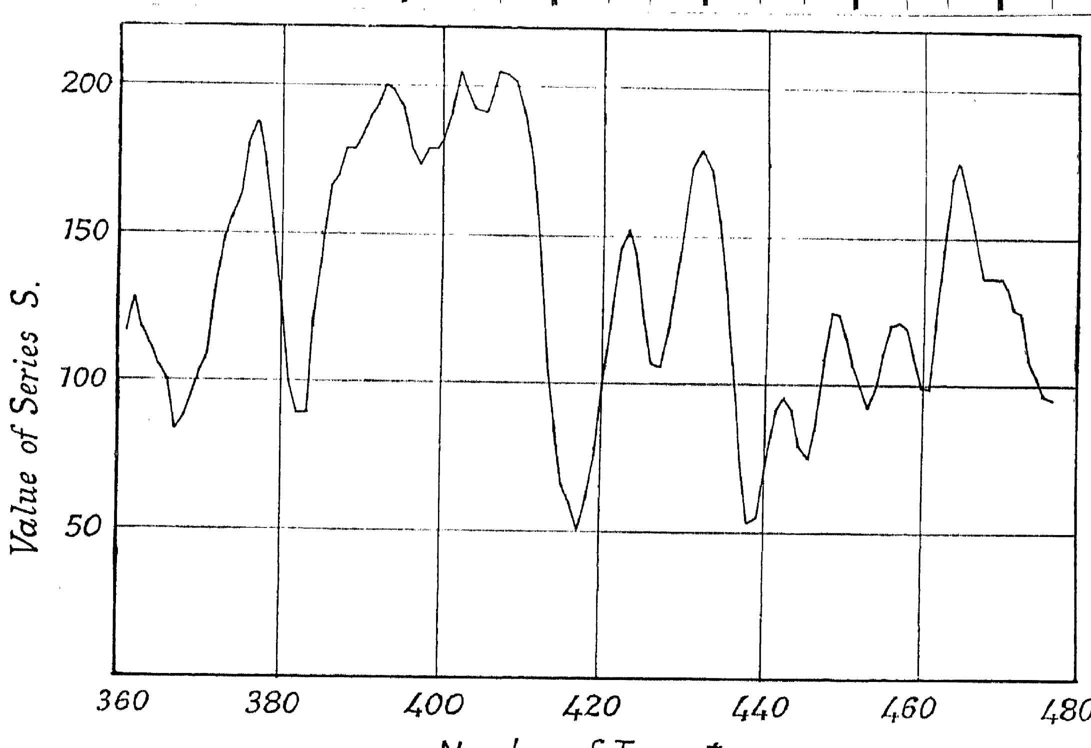


FIG. 29.6.—Graph of the Last 117 Terms of the Series S of Table 29.7.

The observed mean distance is 20·0 units, but this is based on rectangular variation, and we are, perhaps, entitled to expect some difference from normal theory. For rectangular random variables, values distant from the mean occur more frequently, and it is not surprising to find oscillations in the series which do not result in upcrosses.

The number of peaks in the series will be found to be 62, the first at the seventh term, the last at the 466th. Hence the mean distance between peaks is $\frac{459}{61} = 7{\cdot}5$ units. From formula (29.37) we find

$$\cos\theta_1 = \frac{4}{6}, \qquad \theta_1 = 48^\circ\ 11'.$$

Thus the theoretical mean distance is $\frac{360}{48{\cdot}187} = 7{\cdot}5$ units, in good agreement with experiment. It will be observed that several of the distances between peaks are due to very small ripples.

From a number of experiments Dodd (1939*a*) concluded that series generated from rectangular material conformed fairly well to normal theory.

29.29. Let us now examine how the variance of the induced oscillation compares with the variance of the original random series.

The sum of k random elements with variance v has variance kv and its mean has variance v/k. It does not follow that a simple moving average has a variance $1/k$ times that of the random element, because of correlations between successive members in the derived series. If the original series was $\varepsilon_1 \ldots \varepsilon_n$ the derived series is, with weights $a_1 \ldots a_k$,

$$\left.\begin{array}{llll} a_1\,\varepsilon_1 & + a_2\,\varepsilon_2 & + \ldots + a_k\,\varepsilon_k & = \eta_1, \text{ say} \\ a_1\,\varepsilon_2 & + a_2\,\varepsilon_3 & + \ldots + a_k\,\varepsilon_{k+1} & = \eta_2 \\ \cdot & \cdot\quad\cdot & \cdot\quad\cdot\quad\cdot & \cdot \\ a_1\,\varepsilon_{n-k+1} & + a_2\,\varepsilon_{n-k+2} & + \ldots + a_k\,\varepsilon_n & = \eta_{n-k+1} \end{array}\right\} \qquad (29.40)$$

The expected value of the sum of these values is zero since the expected value of ε may be taken to be so. Since there are $n - k + 1$ terms we have for the variance

$$\frac{1}{n-k+1}\,\Sigma\,\eta^2. \qquad (29.41)$$

The expected value of this, since the ε's are independent, is

$$\frac{1}{n-k+1}\,E\,\{\Sigma\,(\eta^2)\,\} = E\,(\eta^2) = (a_1^2 + a_2^2 + \ldots a_k^2)\,v. \qquad (29.42)$$

In particular, if the a's are all equal to $1/k$, the expected value of the variance is v/k. This gives us the *average* reduction in the variance.

If a simple average of extent k is iterated q times the weights are the successive coefficients in

$$\frac{1}{k^q}\,(1 + x + x^2 + \ldots + x^{k-1})^q.$$

The sum of squares of these coefficients is the coefficient of $x^{q\,(k-1)}$ in

$$\frac{1}{k^q}\,(1 + x + x^2 + \ldots + x^{k-1})^{2q} = \frac{(1 - x^k)^{2q}}{k^q\,(1 - x)^{2q}} \qquad (29.43)$$

and this gives the average reduced variance for a simple average of k iterated q times. The following are the values of the reducing factor for some of the values of k and q:—

		q				
		1	2	3	4	5
k	3	0·33	0·23	0·19	0·17	0·15
	4	0·25	0·17	0·14	0·12	0·11
	5	0·20	0·14	0·11	0·10	0·09
	6	0·17	0·11	0·09	0·08	0·07
	7	0·14	0·10	0·08	0·07	0·06

Evidently the result of the first moving average is to generate a series with a much lower variance than that of the original random element, but the second and succeeding iterations do not reduce the variance further to the same extent. In the case $k = 7$ the first averaging reduces the variance to one-seventh, but the next three reduce it only by a further half.

29.30. To apply such results in practice we require an estimate of the variance of the random element in the original series. If this is available we can estimate the variance of the generated series and also, from **29.26**, the mean distance between upcrosses or between peaks. If then our residual series, after the elimination of trend, showed an oscillatory movement with this variance and these mean-distances, within sampling limits, we could not conclude that the oscillatory effect was real. It could have been induced by our method of eliminating trend.

In the present state of knowledge it is not possible to assign permissible limits of sampling variation by relation to standard errors in the usual way. Whether any particular effect is significantly different from the values of the series generated from the random element remains, therefore, a matter of subjective judgment to some extent. The sampling problems involved are formidable, but there does not seem any reason why they should not be capable of explicit solution. This field of study awaits the attention of the theorist.

Example 29.5

For the data of Table 29.3 (sheep population of England and Wales) trend was eliminated by a simple average of nines, the resulting residuals being shown in Table 29.8. A glance at the series suggests some sort of oscillatory effect, since the signs of terms cluster together. By the methods of the next chapter the effect may be brought into greater prominence. The data themselves, however, indicate a mean-distance between upcrosses of about 8 or 9 years, and actual calculation gives a variance of 8474. Can this be due to the operation of our trend-elimination on a random element in the original series?

For the mean distance between upcrosses due to a simple nine-point average we have

$$\cos\theta = \frac{8}{9}, \qquad \theta = 27^\circ\ 16',$$

and the mean distance is $\dfrac{360}{27{\cdot}27} = 13{\cdot}2$ approximately. This is considerably in excess of our observed value, but not sufficiently so to reject outright the possibility we are examining.

Since, however, the variance of residuals is 8474 this must, to have been generated from a random series by a simple average of nines, derive from a random element with

TABLE 29.8

Residual Values of the Sheep Series of Table 29.3 after Elimination of Trend by a Simple Nine-Point Moving Average.

Year.	Residual (10,000).	Year.	Residual (10,000).	Year.	Residual (10,000).
1871	− 176	1893	+ 34	1915	+ 19
72	− 112	94	− 103	16	+ 128
73	+ 50	95	− 104	17	+ 97
74	+ 141	96	− 15	18	+ 69
75	+ 60	97	− 23	19	− 29
76	− 20	98	+ 17	20	− 174
77	+ 12	99	+ 71	21	− 107
78	+ 82	1900	+ 35	22	− 142
79	+ 130	01	+ 16	23	− 109
80	− 14	02	− 27	24	− 23
81	− 166	03	− 32	25	+ 60
82	− 179	04	− 49	26	+ 121
83	− 84	05	− 61	27	+ 94
84	+ 38	06	− 52	28	− 25
85	+ 97	07	− 24	29	− 90
86	+ 8	08	+ 68	30	− 75
87	− 5	09	+ 141	31	+ 72
88	− 105	10	+ 119	32	+ 152
89	− 99	11	+ 66	33	+ 112
90	+ 35	12	− 52	34	− 64
91	+ 159	13	− 117	35	− 87
92	+ 167	14	− 61		

variance 76,266. An estimate of the variance of the random element in the original series, obtained by the variate-difference method which we describe below, was only 350 approximately. Making every allowance for sampling effects, we cannot do otherwise than reject decisively the possibility that the residual oscillation is spurious in the sense of having been induced into the data by the effect of the elimination of trend on a random element.

29.31. We may summarise the foregoing discussion of trend-elimination as follows :—

(*a*) The conception of a trend as a "smooth" or "regular" movement is equivalent to the supposition that trend can be represented, at least locally, by a smooth mathematical function and in particular by a polynomial in the time-variable.

(*b*) Certain series can be treated on lines formally equivalent to regression analysis ; but a more generally applicable procedure is to represent the trend by a moving parabolic arc.

(*c*) The moving arc of best fit in the least-squares sense gives values which are derivable from a moving average of the data. The weights of this average are to some extent at choice, according to the extent of the average and the closeness of fit required in the moving arc.

(*d*) A moving average of extent k sacrifices $(k - 1)$ terms, in the sense that the derived series is $(k - 1)$ terms shorter than the original series. If the series is short it is usually desirable to keep this loss to a minimum, that is, to keep the extent of the average as short as possible.

(*e*) A moving average may distort genuine oscillatory effects, in general exaggerating the shorter variations at the expense of the longer ones, and may induce spurious oscillatory phenomena by its action on random residuals. For harmonic components the effect is minimised by taking the average as simple, with extent equal to the period of the component. For random components the effect is minimised by making the sum of squares of weights in the average a minimum, i.e. by using a simple average.

29.32. In the theory of time-series there are very few rules which can be laid down without a good deal of proviso and caveat. It will be evident from the foregoing that there is no golden rule in trend-fitting which can be applied irrespective of individual circumstances. If we desire to get a close fit to the data we must use a parabola of fairly high order, which involves a moving average with weights which are far from equal. This, however, increases the danger of obscuring the true oscillations in the residuals. In most practical cases it is necessary to strike a balance between conflicting requirements by intuitive judgment as to the appropriate moving average to use.

The Variate-difference Method

29.33. We now proceed to consider the random constituent of a time-series. From the very nature of random variation we cannot expect to derive any formula, however approximate, which will measure the random component directly at any given point of the series. The best we can hope to do is to determine the non-random components and to obtain a random residual which is left unaccounted for by those components ; and even this, as we shall see in the next chapter, is not a very strong hope when oscillations appear in the series.

On certain assumptions, however, we may determine the *variance* of the random component and hence obtain a general idea of its magnitude and importance. Suppose that the systematic part of the series can be represented, at least locally, by a polynomial. Then successive differencing of the series will gradually eliminate the polynomial element but will not reduce the random element correspondingly. As we proceed with the differencing, the random element becomes more and more predominant until finally the systematic component is negligible. Hence we can determine effectively the variance of the random component in the differenced series, and by a simple calculation derive an estimate of that in the original series.

29.34. Consider the differencing of a random series ε_t. We have

$$\Delta\, \varepsilon_t = \varepsilon_{t+1} - \varepsilon_t \qquad (29.44)$$

$$\Delta^r\, \varepsilon_t = \varepsilon_{t+r} - \binom{r}{1}\varepsilon_{t+r-1} + \binom{r}{2}\varepsilon_{t+r-2} + \ldots + (-1)^r\,\varepsilon_t. \qquad (29.45)$$

Without loss of generality we may suppose that the mean value of ε_t is zero, and thus

$$E\,(\Delta^r\,\varepsilon_t) = 0. \qquad (29.46)$$

Hence

$$\begin{aligned}\text{var}\,(\Delta^r\,\varepsilon_t) &= E\,(\Delta^r\,\varepsilon_t)^2 \\ &= E\left\{\varepsilon_{t+r} - \binom{r}{1}\varepsilon_{t+r-1} + \ldots + (-1)^r\,\varepsilon_t\right\}^2 \\ &= E\left\{\varepsilon^2_{t+r} + \binom{r}{1}^2\varepsilon^2_{t+r-1} + \ldots + \varepsilon^2_t\right\} \\ &= v\left\{1 + \binom{r}{1}^2 + \ldots + 1\right\}.\end{aligned}$$

The sum in curly brackets is easily evaluated from the consideration that it is the coefficient of x^r in $(1 + x)^r (x + 1)^r$, that is, equals $\binom{2r}{r}$. Hence

$$\text{var}(\Delta^r \varepsilon_t) = v\binom{2r}{r}. \qquad (29.47)$$

We may then derive an estimate of v by writing

$$v = \frac{\mu_2'(\Delta^r \varepsilon_t)}{\binom{2r}{r}}. \qquad (29.48)$$

It is to be noticed that we use the second moment about zero, not the observed variance of $\Delta^r \varepsilon_t$, since the mean is known to be zero. This shortens the arithmetic to some extent.

The factor $\binom{2r}{r}$ for $r = 1$ to 10 has the following values :—

r	$\binom{2r}{r}$	$1/\binom{2r}{r}$
1	2	0·5
2	6	0·166,667
3	20	0·05
4	70	0·014,285,7
5	252	$0{\cdot}0^2 3{,}968{,}25$
6	924	$0{\cdot}0^2 1{,}082{,}25$
7	3,432	$0{\cdot}0^3{,}291{,}375$
8	12,870	$0{\cdot}0^4 77{,}700{,}1$
9	48,620	$0{\cdot}0^4 20{,}567{,}7$
10	184,756	$0{\cdot}0^5 5{,}412{,}54$

29.35. Basing itself on equation (29.48) the method of variate-differences proceeds as follows : We difference the series once, find the second moment about zero of the resultant and divide by 2 ; we then difference again and find the second moment about zero, dividing in this case by 6 ; and so on. If the successive estimates of v decrease, we continue with the differencing. There will, in general, come a point when they cease decreasing and remain constant within sampling limits (which may be rather wide). At this stage we may suppose that we have eliminated the systematic element in the original series. The final estimate gives us an estimate of the variance of the random element in the original series, and the order of the difference to which we have had to go will give an indication of the degree of the polynomial representing the systematic component.

Example 29.6

Let us apply the variate-difference technique to the series of Table 29.6. We know from the method of constructing the series that the systematic part ought to be completely eliminated after the third differencing, and also that the random part consists of an element with variance 833 approximately. In fact, the random numbers from 1 to N have a variance $(N^2 - 1)/12$ and N in this case is 100. The actual variance of the random element in Table 29.6 is 843.

TABLE 29.9

Differences of the Series u_t of Table 29.6.

t	u_t	Δ^1.	Δ^2.	Δ^3.	Δ^4.	Δ^5.	Δ^6.
1	−96	−6	67	155	279	508	1050
2	−90	−73	−88	−124	−229	−542	−1297
3	−17	15	36	105	313	755	1524
4	−32	−21	−69	−208	−442	−769	−1141
5	−11	48	139	234	327	372	271
6	−59	−91	−95	−93	−45	101	361
7	32	4	−2	−48	−146	−260	−229
8	28	6	46	98	114	−31	−625
9	22	−40	−52	−16	145	594	1661
10	62	12	−36	−161	−449	−1067	−2252
11	50	48	125	288	618	1185	1978
12	2	−77	−163	−330	−567	−793	−876
13	79	86	167	237	226	83	−159
14	−7	−81	−70	11	143	242	137
15	74	−11	−81	−132	−99	105	551
16	85	70	51	−33	−204	−446	−655
17	15	19	84	171	242	209	−64
18	−4	−65	−87	−71	33	273	690
19	61	22	−16	−104	−240	−417	−629
20	39	38	88	136	177	212	216
21	1	−50	−48	−41	−35	−4	175
22	51	−2	−7	−6	−31	−179	−650
23	53	5	−1	25	148	471	1110
24	48	6	−26	−123	−323	−639	−975
25	42	32	97	200	316	336	41
26	10	−65	−103	−116	−20	295	925
27	75	38	13	−96	−315	−630	−965
28	37	25	109	219	315	335	207
29	12	−84	−110	−96	−20	128	316
30	96	26	−14	−76	−148	−188	−32
31	70	40	62	72	40	−156	−798
32	30	−22	−10	32	196	642	1597
33	52	−12	−42	−164	−446	−955	−1719
34	64	30	122	282	509	764	950
35	34	−92	−160	−227	−255	−186	141
36	126	68	67	28	−69	−327	−991
37	58	1	39	97	258	664	1515
38	57	−38	−58	−161	−406	−851	−1492
39	95	20	103	245	445	641	707
40	75	−83	−142	−200	−196	−66	281
41	158	59	58	−4	−130	−347	−685
42	99	1	62	126	217	338	509
43	98	−61	−64	−91	−121	−171	−314
44	159	3	27	30	50	143	432
45	156	−24	−3	−20	−93	−289	−745
46	180	−21	17	73	196	456	...
47	201	−38	−56	−123	−260	...	...
48	239	18	67	137	...	...	...
49	221	−49	−70	...	...	...	...
50	270	21	...	...	...	...	...
51	249	...	...	...	...	...	...

Table 29.9 shows the series and the differences up to Δ^6. For the sums of squares in the various columns S_j corresponding to Δ^j, we find—

$$\begin{aligned} S_1 &= 107{,}541 \\ S_2 &= 318{,}115 \\ S_3 &= 1{,}033{,}513 \\ S_4 &= 3{,}445{,}308 \\ S_5 &= 11{,}720{,}069 \\ S_6 &= 40{,}548{,}844 \end{aligned}$$

To obtain second moments we divide by $51 - j$ and then, to obtain the estimate of v, by $\binom{2j}{j}$. We find the following :—

j	Estimate.
1	1075·41
2	1082·02
3	1076·58
4	1047·21
5	1011·05
6	975·20

Curiously enough, the estimate for $j = 2$ is higher than that for $j = 1$ and there is little difference between the various estimates. In the ordinary way we should have concluded that the systematic component was adequately represented by a polynomial of order 1, that is to say a straight line, and that the residual random element had a variance of about 1000.

The reader must not be surprised to find discrepancies of this kind between theory and experiment in short series ; and the discrepancy is not, in fact, as big as it seems. The variance of the original series is 6272·61. The mean square of the first difference, divided by 2, is 1075·41, so that about five-sixths of the variance has been eliminated by the first differencing, and the method indicates, quite correctly, that the greater part of the systematic element is linear. The random element is rather large compared with the non-linear systematic terms, and the latter have got caught up in it—the series is too short for the variate-difference method to disentangle them. Consider, for instance, the cubic term $\frac{1}{100}(t - 26)^3$. In the original series this varies in value from $-156{\cdot}25$ to $+156{\cdot}25$. First differences reduce it to $\dfrac{3}{100}(t - 26)^2$, varying from 18·75 through zero to 18·75, whereas the random element is increased in range from 0 to 198. Already the systematic term is being swamped by the random element, and a slight degree of accidental correlation between the two can easily account for the increase in the mean-square of second differences.

The matter may be put in a slightly different way. Suppose that, relying on the variate-difference method, we regarded the data as represented by a linear equation plus a random residual. If we fitted a straight line by least squares and examined the residuals, we should probably find very little evidence of departure from randomness. This representation would differ from the mode of construction of the series, but it would be a *possible* method of construction. Only the failure of the representation to conform to further terms of the series would reveal its weakness.

29.36. The variate-difference method thus provides a kind of lower limit to the degree of the polynomial which will represent a series locally or generally. There remains for consideration the question as to what sort of differences between successive estimates of v can be regarded as chance effects, in order to decide when the value has reached a stationary level. The sum of squares S_j is a constant factor times the second moment, but as its members are correlated among themselves we cannot use the variance of the second moment to test its significance. Further, S_j and S_{j+1} are correlated. We proceed to derive the sampling variance of their difference, the somewhat complicated formulae being due to Anderson (1914).

29.37. Write

$$b_j = \binom{r}{j}. \qquad (29.49)$$

Then we have, as in (29.42),

$$\frac{1}{\binom{2r}{r}} E\,(\Delta^r u)^2 = \frac{(b_0^2 + b_1^2 + \ldots)\,\mu_2}{(b_0^2 + b_1^2 + \ldots)} = \mu_2, \qquad (29.50)$$

where μ_2 is the variance of u. Further

$$\begin{aligned} E\,(\Delta^r u)^4 = E\,[\,&\{b_0 u_{r+1} - b_1 u_r + b_2 u_{r-1} - \ldots + (-1)^r b_r u_1\}^2 \\ &+ \{b_0 u_{r+2} - b_1 u_{r+1} + b_2 u_r - \ldots + (-1)^r b_r u_2\}^2 \\ &+ \quad \ldots \\ &+ \{b_0 u_n - b_1 u_{n-1} + b_2 u_{n-2} - \ldots + (-1)^r b_r u_{n-r}\}^2]^2. \end{aligned} \qquad (29.51)$$

Consider first of all the terms in this which result in fourth powers of u. They will derive from

$$\begin{aligned} E\,&\{b_0^2 u_{r+1}^2 + b_1^2 u_r^2 + \ldots + b_r^2 u_1^2 + b_0^2 u_{r+2}^2 + b_1^2 u_{r+1}^2 + \ldots + b_r^2 u_2^2 + \ldots \\ &+ b_0^2 u_n^2 + b_1^2 u_{n-1}^2 + \ldots + b_r^2 u_{n-r}^2\}^2 \\ = E\,&\{b_0^2 (u_n^2 + u_1^2) + (b_0^2 + b_1^2)(u_{n-1}^2 + u_2^2) + (b_0^2 + b_1^2 + b_2^2)(u_{n-2}^2 + u_3^2) + \ldots \\ &+ (b_0^2 + b_1^2 + \ldots b_{r-1}^2)(u_{n-r+1}^2 + u_r^2) + (b_0^2 + b_1^2 + \ldots + b_r^2) \\ &(u_{n-r}^2 + u_{n-r-1}^2 + \ldots + u_{r+1}^2)\}^2. \end{aligned} \qquad (29.52)$$

Writing now

$$B_0^2 = (b_0^2)^2 + (b_0^2 + b_1^2)^2 + \ldots + (b_0^2 + b_1^2 + \ldots + b_{r-1}^2)^2 \qquad (29.53)$$

$$A_0^2 = (b_0^2 + b_1^2 + \ldots + b_r^2)^2 = \binom{2r}{r}^2. \qquad (29.54)$$

we see that the term in $E\,(u^4)$ is

$$\{A_0^2\,(n - 2r) + 2B_0^2\}\,E\,(u^4). \qquad (29.55)$$

The only other term appearing from (29.51) will be of type $E\,(u_l^2 u_m^2)$, $l \neq m$. If the reader will write out the expansion of (29.51) he will find that the coefficients are expressible in terms of

$$A_j^2 = (b_0 b_j + b_1 b_{j+1} + \ldots + b_{r-j} b_r)^2 = \binom{2r}{r-j}^2 \qquad (29.56)$$

and

$$B_j^2 = (b_0 b_j)^2 + (b_0 b_j + b_1 b_{j+1})^2 + \ldots + (b_0 b_j + b_1 b_{j+1} + \ldots + b_{r-j-1} b_{r-1})^2. \qquad (29.57)$$

The expression for $E(\Delta^r u)^4$ reduces to—

$$(n-2r)A_0^2 E(u^4) + 4\{(n-2r+1)A_1^2 + (n-2r+2)A_2^2 + \dots + A_r^2(n-2r+r)\}E(u_l^2 u_m^2) + 2B_0^2 E(u^4) + 8\{B_1^2 + B_2^2 + \dots + B_{r-1}^2 + B_r^2\}E(u_l^2 u_m^2). \quad (29.58)$$

Substituting μ_4 for $E(u^4)$ and μ_2^2 for $E(u_l^2 u_m^2)$, dividing by $(n-r)^2\binom{2r}{r}^2$ and subtracting μ_2^2, we find the sampling variance of the estimate of v. The expression can, however, be simplified to some extent. Putting

$$T_r = \sum_{j=0}^{r-1}\binom{r}{j}^2\binom{r}{j+1}^2 + 2\sum_{j=0}^{r-2}\binom{r}{j}^2\binom{r}{j+2}^2 + 3\sum_{j=0}^{r-3}\binom{r}{j}^2\binom{r}{j+3}^2 + \dots + r\binom{r}{0}^2\binom{r}{r}^2 \quad (29.59)$$

we find, after lengthy algebraic rearrangement,

$$\operatorname{var}\frac{S_r}{(n-r)\binom{2r}{r}} = \frac{\mu_4 - 3\mu_2^2}{n-r}\left\{1 - \frac{2T_r}{(n-r)\binom{2r}{r}^2}\right\} + \frac{2\mu_2^2}{n-r}\left\{\frac{\binom{4r}{2r}}{\binom{2r}{r}^2} - \frac{r}{2(n-r)}\right\}, \quad r \leqslant \tfrac{1}{2}n. \quad (29.60)$$

If terms of order $(n-r)^{-2}$ can be neglected, this reduces to

$$\frac{\mu_4 - 3\mu_2^2}{n-r} + \frac{\binom{4r}{2r}}{\binom{2r}{r}^2}\frac{2\mu_2^2}{n-r}, \quad (29.61)$$

or, using the Stirling approximation to factorials,

$$\frac{1}{n-r}\{\mu_4 - 3\mu_2^2 + \mu_2^2\sqrt{(2r\pi)}\}, \quad (29.62)$$

which is a fair approximation to (29.61), being within 3 per cent. for r as low as 6.

When the population of values of u is normal, $\mu_4 - 3\mu_2^2$ vanishes and the formula simplifies accordingly.

29.38. In a similar way it may be shown that

$$\operatorname{cov}\left\{\frac{S_r}{(n-r)\binom{2r}{r}}, \frac{S_{r+1}}{(n-r-1)\binom{2r+2}{r+1}}\right\}$$

$$= \frac{\mu_4 - 3\mu_2^2}{n-r}\left\{1 - \frac{2T_r'}{\binom{2r}{r}\binom{2r+2}{r+1}(n-r-1)}\right\} + \frac{2\mu_2^2}{n-r}\left\{\frac{\binom{4r+1}{2r}}{\binom{2r}{r}\binom{2r+2}{r+1}}\frac{2n-2r-1}{n-r-1} - \frac{r+1}{2(n-r-1)}\right\}. \quad (29.63)$$

where

$$T'_r = \sum_{j=0}^{r-1} \binom{r}{j}^2 \binom{r+1}{j+2}^2 + 2\sum_{j=0}^{r-2} \binom{r}{j}^2 \binom{r+1}{j+3}^2 + \ldots + r\binom{r}{0}^2 \binom{r+1}{r+1}^2.$$

From (29.60) and (29.63) we can determine the variance of the difference of

$$\frac{S_r}{(n-r)\binom{2r}{r}} \quad \text{and} \quad \frac{S_{r+1}}{(n-r-1)\binom{2r+2}{r+1}}.$$

The general formula is complicated, but for normal variation, large n and $r \geqslant 6$ we have, analogously to (29.62),

$$\operatorname{var}\left\{\frac{S_r}{(n-r)\binom{2r}{r}} - \frac{S_{r+1}}{(n-r-1)\binom{2r+2}{r+1}}\right\}$$

$$= \frac{(3r+1)\sqrt{(2\pi r)}}{2\,(2r+1)^3\,(n-r-1)}\left\{\frac{S_r}{(n-r)\binom{2r}{r}}\right\}^2. \qquad (29.64)$$

The arithmetic application of the formulae has been facilitated by the preparation of tables of the constants involved. Reference may be made to Tintner (1940) who gives tables prepared by himself, Anderson and Zaycoff.

Example 29.7

For the data of Table 29.3 (sheep population) an application of the variate-difference method up to the tenth difference gave the following results :—

r	$S_r\Big/\binom{2r}{r}(n-r)$
1	3468
2	1442
3	854
4	629
5	518
6	448
7	401
8	371
9	357
10	347

The values here are falling steadily from $r = 1$ to $r = 10$, but very slightly towards the end. From (29.64) for $r = 6$ we have for the variance of the difference, 80·7 approximately and for $r = 10$, 25·8 approximately. It appears that the reduction in variance at $r = 10$ is losing significance, and that a moving arc of degree 10 would be sufficient to eliminate the systematic component. It does not, of course, follow that the trend-line must be of this degree, for we may not want to eliminate the oscillatory movements in the trend-line.

29.39. The variate-difference method will clearly not eliminate systematic effects such as periodic terms with very short period. Consider, for instance, the series 1, − 1, 1, − 1, etc. The first differences give us a series 2, − 2, 2, − 2, etc., second differences

4, $-$ 4, 4, $-$ 4, etc., and so on. The variance of the series of rth differences is, neglecting effects due to the shortness of the series, 2^{2r} times that of the original, and the quotient when this is divided by $\binom{2r}{r}$ tends to

$$\frac{2^{2r}\,(r\,!)^2}{(2r\,!)} \rightarrow \sqrt{\pi r}$$

and so increases without limit. In such a case we cannot obtain an estimate of the variance of any random element which may be present.

NOTES AND REFERENCES

References to the fitting of polynomials are given at the end of Chapter 22. For the moving average see Whittaker and Robinson's *Calculus of Observations* and the books by Macaulay (1931) and Sasuly (1934).

Attempts have been made to use trend-lines for purposes of forecasting, and even to measure the standard error of a forecast—see Schultz (1930) and a discussion in Davis (1941). The methods proposed appear to me theoretically unsound and in practice they lead as a rule to such wide limits of error as to be of doubtful value ; but this is a personal opinion and the less sceptical reader may care to consult Davis's book and to follow up the references given therein.

For the effect of moving averages on random variables see Yule (1921) and Slutzky (1937*b*), the latter being an English version of a paper published in Russian many years earlier. See also Dodd (1939*a*, 1941*a*). Slutzky proves an interesting theorem—the theorem of the sinusoidal limit—to the effect that repeated moving averages of certain kinds applied to random series generate a sine-curve.

For the variate-difference method see the book by Tintner (1940), a very thorough practical account with useful tables. The more important earlier memoirs are those by Anderson (1914, 1923, 1926), "Student" (1914), Morant (1921), and K. Pearson and Cave (1914).

EXERCISES

29.1. Show that in the formulae of equation (29.7) and similar formulae of higher orders the sum of the weights is unity.

29.2. By evaluating the solutions of (29.5) determinantally show that a parabolic curve of second or third order giving a graduation

$$a_t\,u_{-t} + a_{(t-1)}\,u_{(t-1)} \;.\;.\;. + a_0\,u_0 + \;.\;.\;. + a_t\,u_t$$

has

$$a_j = 3\,\frac{3n^2 + (3n-1) - 5j^2}{(2n-1)\,(2n+1)\,(2n+3)}.$$

29.3. Show that the weights in the Spencer 21-point formula are

$$\frac{1}{350}[-1,\ -3,\ -5,\ -5,\ -2,\ 6,\ 18,\ 33,\ 47,\ 57,\ \mathbf{60},\ .\ .\ .]$$

and that if it is applied to a random series the variance of the resultant is about one-seventh

of the original series—about the same reduction as would be given by a simple moving average of sevens.

29.4. Show that Macaulay's 43-point formula,

$$\frac{1}{960}[12][8][5]^2\left[\frac{7}{10}, -1, 0, 0, 0, 0, 0, 0, \mathbf{1}, \ldots\right],$$

has weights

$$\frac{1}{9600}[7, 18, 30, 40, 45, 28, -8, -60, -122, -178, -205, -190, -127,$$
$$-6, 163, 360, 562, 760, 928, 1050, 1127, \mathbf{1156}, \ldots]$$

and that it reduces the variance of a random series about as much as a simple average of nines.

29.5. Take a random series of, say, 200 terms and determine "trends" by moving averages $\frac{1}{9}[9]$, $\frac{1}{81}[9]^2$ and $\frac{1}{729}[9]^3$. Compare the mean distances between peaks and upcrosses with the theoretical values based on normal theory.

29.6. If ε_t is a random series, show that the correlation between successive members of $\Delta^k \varepsilon_t$ for long series is $-\frac{k}{k+1}$ and hence tends to -1 as k increases. Hence show that the signs of successive terms in $\Delta^k u_t$ tend to alternate, where u_t is the sum of a random element and a systematic element representable by a polynomial; and verify by reference to Table 29.9.

29.7. By eliminating δ^2 from (29.19) show that, for a cubic curve, an accurate trend-line is given by

$$\frac{1}{h^2-k^2}\left\{\frac{h^2-1}{k}[k] - \frac{k^2-1}{h}[h]\right\}$$

and generalise this result.

(Cf. J. A. Higham, *J. Inst. Act.* (1882-5), **23**, 335; **25**, 15, 245.)

CHAPTER 30

TIME-SERIES—(2)

30.1. The present chapter is devoted to a discussion of oscillatory effects in time-series. We shall suppose that our series is *stationary*, i.e. has no trend, either because the original data contained none or because trend has been removed by one of the methods described in the last chapter. Our typical series will then fluctuate round some constant value which we may usually, without loss of generality, take to be zero. We shall assume that there is a prior possibility that part of the variation at least is random. This, indeed,

TABLE 30.1

Trend-free Wheat-Price Index (European Prices) compiled by Sir William Beveridge for the Years 1500–1869.

(From Beveridge, 1921.)

Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.	Year.	Index.
1500	106	1537	73	1574	113	1611	100	1648	122	1685	74	1722	91	1759	91	1796	95	1833	80
01	118	38	86	75	89	12	99	49	134	86	75	23	94	60	88	97	84	34	78
02	124	39	74	76	87	13	100	50	119	87	66	24	110	61	100	98	87	35	82
03	94	40	74	77	87	14	94	51	136	88	62	25	111	62	97	99	120	36	88
04	82	41	76	78	79	15	88	52	102	89	76	26	103	63	88	1800	139	37	102
05	88	42	80	79	90	16	92	53	72	90	79	27	94	64	95	01	117	38	117
06	87	43	96	80	90	17	100	54	63	91	97	28	101	65	101	02	105	39	107
07	88	44	112	81	87	18	82	55	76	92	134	29	90	66	106	03	94	40	95
08	88	45	144	82	83	19	73	56	75	93	169	30	96	67	113	04	125	41	101
09	68	46	80	83	85	20	81	57	77	94	111	31	80	68	108	05	114	42	92
10	98	47	54	84	76	21	99	58	103	95	109	32	76	69	108	06	98	43	88
11	115	48	69	85	110	22	124	59	104	96	111	33	84	70	131	07	93	44	92
12	135	49	100	86	161	23	106	60	120	97	128	34	91	71	136	08	94	45	115
13	104	50	103	87	97	24	106	61	167	98	163	35	94	72	119	09	94	46	139
14	96	51	129	88	84	25	121	62	126	99	137	36	101	73	106	10	104	47	90
15	110	52	100	89	106	26	105	63	108	1700	99	37	93	74	105	11	140	48	80
16	107	53	90	90	111	27	84	64	91	01	85	38	91	75	88	12	121	49	74
17	97	54	100	91	97	28	97	65	85	02	72	39	122	76	84	13	96	50	78
18	75	55	123	92	108	29	109	66	73	03	88	40	159	77	94	14	96	51	86
19	86	56	156	93	100	30	148	67	74	04	77	'41	110	78	87	15	130	52	105
20	111	57	71	94	119	31	114	68	80	05	66	42	90	79	79	16	178	53	138
21	125	58	71	95	131	32	108	69	74	06	64	43	81	80	87	17	126	54	141
22	78	59	81	96	143	33	97	70	78	07	69	44	84	81	88	18	94	55	138
23	86	60	84	97	138	34	92	71	83	08	125	45	102	82	94	19	86	56	107
24	102	61	97	98	112	35	97	72	84	09	175	46	102	83	94	20	84	57	82
25	71	62	105	99	99	36	98	73	106	10	108	47	100	84	92	21	76	58	81
26	81	63	90	1600	97	37	105	74	134	11	103	48	109	85	85	22	77	59	97
27	129	64	78	01	80	38	97	75	122	12	115	49	104	86	84	23	71	60	116
28	130	65	112	02	90	39	93	76	102	13	134	50	90	87	93	24	71	61	107
29	129	66	100	03	90	40	99	77	107	14	108	51	99	88	108	25	69	62	92
30	125	67	86	04	80	41	99	78	115	15	90	52	95	89	108	26	82	63	79
31	139	68	77	05	77	42	107	79	113	16	89	53	90	90	86	27	93	64	81
32	97	69	80	06	81	43	106	80	104	17	89	54	80	91	78	28	114	65	94
33	90	70	93	07	98	44	96	81	92	18	94	55	85	92	87	29	103	66	119
34	76	71	112	08	115	45	82	82	84	19	107	56	117	93	85	30	110	67	118
35	102	72	131	09	94	46	88	83	86	20	89	57	112	94	103	31	105	68	93
36	100	73	158	10	93	47	116	84	101	21	79	58	95	95	130	32	82	69	102

is necessary if our results are to have any practical application, for most of the series encountered in practice have some element of irregularity, however small.

30.2. Four examples of the type of series under consideration have already occurred. The table of Example 21.11 (page 126) gives the deviations from a simple nine-year moving average of the yields of potatoes in tenths of tons per acre in England and Wales for the years 1888–1935. Table 29.1 (Fig. 29.1) gives the annual yields of barley in cwts. per acre in England and Wales for 1884–1939, no nine-year elimination of trend having been carried out in this case. Table 29.4 (Fig. 29.4) gives rainfall data at London over the century 1813–1912. Table 29.5 (Fig. 29.5) gives egg-production per laying hen in the U.S.A.

TABLE 30.2

Marriage Rate in England and Wales: Deviation from a Simple 11-Year Moving Average for the Years 1843–1896.

Units 1 in 10,000.

Year.	Marriage Rate.	Year.	Marriage Rate.	Year.	Marriage Rate.
1843	− 6	1861	− 5	1879	− 12
44	1	62	− 7	80	− 5
45	12	63	1	81	0
46	10	64	6	82	5
47	− 6	65	8	83	7
48	− 8	66	9	84	3
49	− 6	67	− 2	85	− 4
50	3	68	− 8	86	− 8
51	4	69	− 10	87	− 6
52	7	70	− 7	88	− 5
53	11	71	0	89	1
54	3	72	8	90	6
55	− 8	73	12	91	6
56	− 2	74	7	92	2
57	− 3	75	5	93	− 6
58	− 7	76	4	94	− 5
59	3	77	− 3	95	− 6
60	4	78	− 6	96	1

Tables 30.1 and 30.2 give two further examples. The first is a famous series of trend-free wheat-price indices compiled by Sir William Beveridge and extending over 370 years, a phenomenal length of time for economic series. The second is the deviation from a simple 11-year moving average of marriage rates for the years 1843–1896.

Oscillation and Cycle

30.3. We will now attempt to define more closely the sense in which we use the words "oscillation" and "cycle". It is particularly important to exercise great care in the use of an accurate nomenclature because a great deal of the literature on this subject suffers from confusion due to loose wording.

By a *cyclical* component of a time-series we shall mean one which is a strictly periodic function of the time, that is to say, for which there exists a *period* ω such that

$$u_t = u_{t+\omega} = u_{t+2\omega} = \dots = u_{t+k\omega} = \dots \qquad (30.1)$$

whatever the value of t. The periodic functions which we shall consider in particular are the sine and cosine functions. If the series can be represented as the sum of a cyclical component and a random constituent, or by a cyclical component alone, we may speak of it as a cyclical series.

30.4. If the series is not random it must move with more or less regularity about the mean value, and we shall then speak of it as *oscillatory*. The oscillatory movement may be in part due to random elements but must not be entirely so. A cyclical series is oscillatory, but an oscillatory series is not necessarily cyclical.

An oscillatory movement may be the sum of two or more cyclical components. Consider, for instance, the sum of two periodic terms

$$u_t = \sin\frac{2\pi t}{\omega_1} + \sin\frac{2\pi t}{\omega_2}.$$

If ω_1 and ω_2 are commensurable there will be numbers, and in particular a smallest number ω, which is an exact multiple of both of them. This is clearly a period of the series. But if ω_1 and ω_2 are not commensurable there will be no period of this kind and the sum will be oscillatory but not cyclical.

30.5. It may be felt by the reader that we could reasonably extend the use of the word "cyclical" to cover series which are the sum of cyclical terms; but the danger of doing so is that within certain limits any series can be represented as a sum of harmonic terms, even if it is not itself oscillatory, in virtue of Fourier's theorem. Admittedly such a representation, to be exact, must in general consist of an infinite series of terms and is valid only in a certain range, but in practice a comparatively small number of terms often gives quite a good approximation. We do not call a function a polynomial because it can be expanded in powers of the variable by Taylor's theorem; and correspondingly we shall not call it cyclical because it can be expanded as a sum of harmonic terms by Fourier's theorem. On the whole it seems safer to avoid the word "cyclical" for series which consist of a finite number of cyclical terms.

30.6. For our present purposes the main significance of the distinction we are attempting to make is that in a cyclical series the maxima and minima, apart from disturbances due to the superposition of a random element, occur at equal intervals of time and are therefore predictable for a long way into the future—for so long, in fact, as the constitution of the system remains unchanged. In oscillatory series, on the other hand, the distances from peak to peak, trough to trough or upcross to upcross, are not equal, but vary very considerably. Similarly, in the oscillatory series the amplitudes of the movements may vary very substantially, whereas in a cyclical series they should be constant (again, except in so far as superposed random elements disturb them).

30.7. Now the time-series observed in practice are very rarely cyclical as we have defined the term. The only case among those cited at the beginning of the chapter in which there appears to be any cyclical movement is that of egg-production per hen in Table 29.5. The far more usual case is that of varying amplitude and period from peak to peak or upcross

to upcross. We shall therefore begin our study of oscillatory movements by considering the kinds of scheme which can give rise to the observed phenomena; and then we shall examine methods of deciding which of the possible schemes should be chosen as the hypothetical representation in particular cases.

Tests for Randomness

30.8. The first stage, when confronted with a fluctuating stationary series, is to examine whether the fluctuations are purely random. Tests of randomness are easy to find, and in fact the random series is the happy hunting-ground of the worker whose interests lie mainly in the mathematics of the direct theory of probability. We have considered some tests which are appropriate to the study of oscillatory movement in **21.43** to **21.46**. Others which have gained popularity are based on the distribution of "runs" and on the correlation between successive members of the series. The reader will have no difficulty in composing others. All these tests are based on the non-parametric case, so that the alternative hypotheses are not usually brought specifically into view. We cannot therefore apply the general theory of Chapters 26 and 27 to determine "best" tests, and in the present state of knowledge are forced to be content with less definite ideas. So far as ease of application goes, the tests of **21.43** and **21.44** seem to have decided advantages, though they may be somewhat insensitive. The method of serial correlation, to which we refer below, gives a useful alternative in doubtful cases. In the sequel we shall suppose that before proceeding to search for systematic movements we have satisfied ourselves by one or more of these tests that such movements exist.

30.9. We shall consider three schemes which can account for the typical oscillatory movement usually observed.

(*a*) *Moving Averages.*—We have already seen in Chapter 29 that a moving average of a purely random element can generate an oscillatory series with all the required properties of varying amplitude and mean distances—the Slutzky-Yule effect (**29.25**). Fig. 29.6 illustrates the kind of oscillation which may arise. It is at least possible that some of the observed oscillations in time-series may be generated in this way; and in fact Slutzky (1936) has given an interesting example in which a part of his series generated by the moving average happens to agree very closely with an observed series.

(*b*) *Sums of Cyclical Components.*—We may attempt, by Fourier analysis or the more general harmonic analysis, to represent the oscillations as the sum of a number of cyclical components. This is the classical approach.

(*c*) *Autoregression Equations.*—If a series is constructed by the recurrence formula

$$u_{t+1} = f(u_t, u_{t-1}, \ldots u_{t-k}) + \varepsilon_{t+1}, \qquad (30.2)$$

where f is a mathematical function and ε a "disturbance" function which may be a random variable, then under certain conditions the generated series is of the required type. We shall consider in particular the series

$$u_{t+2} = -au_{t+1} - bu_t + \varepsilon_{t+2}, \qquad (30.3)$$

where a and b are constants and ε is random.

Table 30.3 (Fig. 30.1) shows a series of type (*b*) in the simplest case where only one cyclical component is involved, together with a random residual. Table 30.4 (Fig. 30.2) shows an autoregressive series constructed from random numbers by the formula

$$u_{t+2} = 1{\cdot}1\, u_{t+1} - 0{\cdot}5\, u_t + \varepsilon_{t+2}. \qquad (30.4)$$

TABLE 30.3

Values of the Series $u_t = 10 \sin \frac{\pi t}{5} + \varepsilon_t$ *where* ε_t *is a Rectangular Random Variable with Range* -5 *to* $+5$*, rounded off to Nearest Unit.*

Number of Term.	Series.	Number of Term.	Series.	Number of Term.	Series.
1	3	21	11	41	5
2	8	22	13	42	12
3	6	23	10	43	7
4	2	24	6	44	5
5	− 4	25	− 5	45	3
6	− 7	26	− 8	46	− 2
7	− 9	27	− 12	47	− 12
8	− 9	28	− 10	48	− 12
9	− 10	29	− 7	49	− 8
10	− 1	30	0	50	− 1
11	8	31	1	51	11
12	7	32	8	52	13
13	6	33	13	53	12
14	4	34	7	54	7
15	− 3	35	4	55	5
16	− 10	36	− 9	56	− 1
17	− 11	37	− 9	57	− 6
18	− 15	38	− 6	58	− 14
19	− 4	39	− 4	59	− 8
20	4	40	− 2	60	1

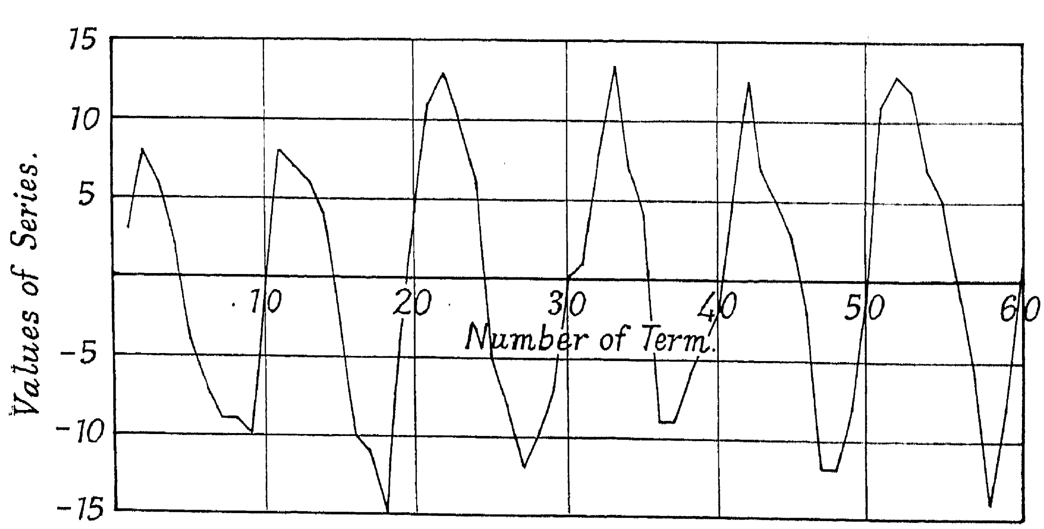


FIG. 30.1.—Graph of the Values of Table 30.3.

TABLE 30.4

Values of Series $u_{t+2} = 1{\cdot}1\ u_{t+1} - 0{\cdot}5\ u_t + \varepsilon_{t+2}$ *where* ε_{t+2} *is a Rectangular Random Variable with Range* $-9{\cdot}5$ *to* $9{\cdot}5$*, rounded off to Nearest Unit.*

Number of Term.	Value of Series.	Number of Term.	Value of Series.	Number of Term.	Value of Series.
1	7	23	− 4	45	− 13
2	6	24	− 5	46	1
3	− 6	25	− 9	47	6
4	− 4	26	− 4	48	4
5	3	27	− 4	49	11
6	− 4	28	3	50	15
7	− 5	29	9	51	9
8	− 1	30	4	52	8
9	10	31	− 8	53	4
10	10	32	− 6	54	− 1
11	6	33	− 3	55	4
12	− 4	34	− 2	56	7
13	− 4	35	0	57	11
14	− 7	36	− 1	58	0
15	− 2	37	− 3	59	1
16	6	38	3	60	0
17	17	39	− 1	61	− 5
18	24	40	− 8	62	− 11
19	17	41	− 3	63	− 8
20	4	42	− 8	64	− 3
21	1	43	− 10	65	5
22	− 5	44	− 16		

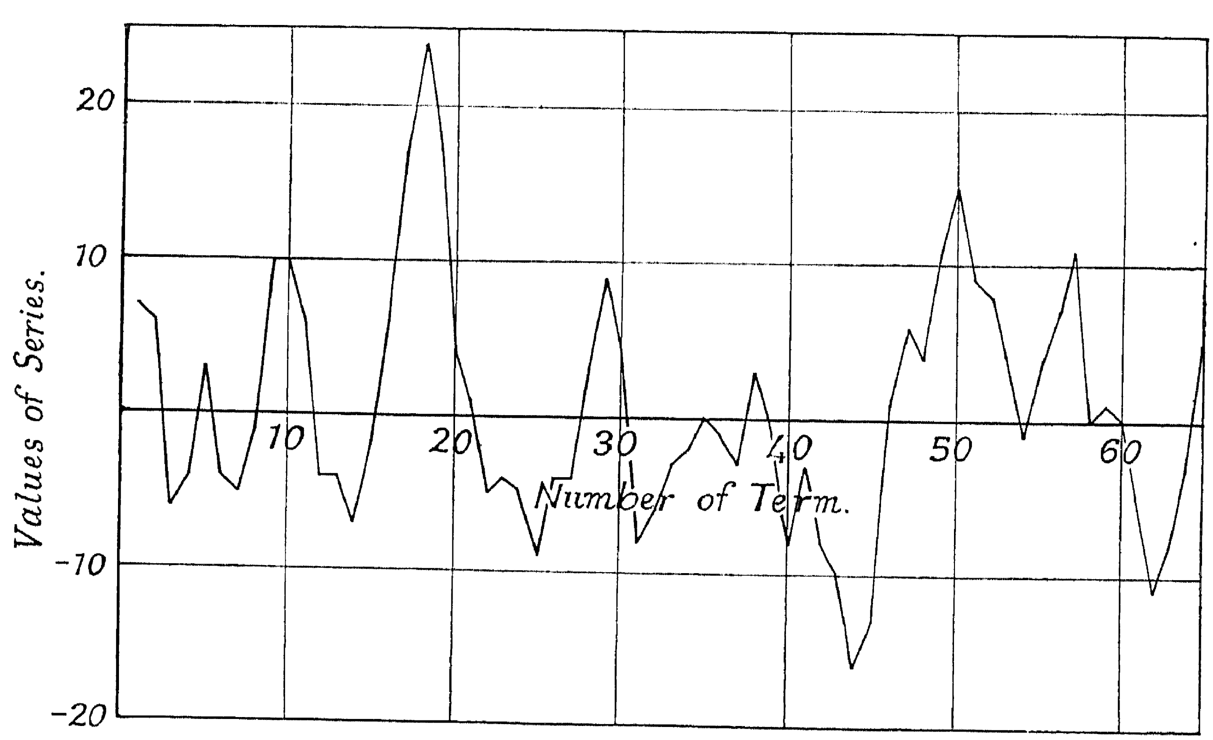


FIG. 30.2.—Graph of the Values of Table 30.4.

30.10. It is quite possible that theoretical reasons may suggest other schemes for study as the subject progresses. For instance, we might wish to consider series defined by differential equations, on the analogy of the similar equations determining oscillations in physical phenomena such as vibrating strings or electrical discharges. Something has, in fact, already been done in this direction. We shall, however, confine our attention to the three schemes indicated above, and particularly the second and third.

30.11. On the face of it, an observed series exhibiting the typical movements in amplitude and period might be due to any one of the three schemes or even to a combination of them. We require, in the first instance, some objective criterion for deciding which of them is applicable in particular cases. Inspection of the primary data, though useful, is quite an unreliable guide in making a decision on this point, particularly if the series is short. Experience seems to indicate that few things are more likely to mislead in the theory of oscillatory series than attempts to determine the nature of the oscillatory movement by mere contemplation of the series itself; and yet this is the method, if one can dignify it by such a term, which has perhaps been most widely used in the past.

Serial Correlation

30.12. Suppose our series of values is $u_1 \ldots u_n$. Let us form the product–moment correlation coefficient between successive terms, i.e.

$$r_1 = \frac{\operatorname{cov}(u_j, u_{j+1})}{(\operatorname{var} u_j \operatorname{var} u_{j+1})^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.5)$$

There will be $(n - 1)$ pairs entering into the correlation, and the variances of u_j and u_{j+1} differ only in the fact that the first relates to the terms $u_1, u_2, \ldots u_{n-1}$ and the second to the terms $u_2, u_3, \ldots u_n$. The coefficient r_1 is called the *serial correlation coefficient* of the first order, or more briefly the first serial correlation.*

More generally, let us define a coefficient of order k:

$$r_k = \frac{\operatorname{cov}(u_j, u_{j+k})}{(\operatorname{var} u_j \operatorname{var} u_{j+k})^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.6)$$

$$= \frac{\dfrac{1}{n-k}\displaystyle\sum_{j=1}^{n-k}(u_j\, u_{j+k}) - \frac{1}{(n-k)^2}\left(\sum_{j=1}^{n-k} u_j\right)\left(\sum_{j=1}^{n-k} u_{j+k}\right)}{\left\{\dfrac{1}{n-k}\displaystyle\sum_{j=1}^{n-k} u_j^2 - \frac{1}{(n-k)^2}\left(\sum_{j=1}^{n-k} u_j\right)^2\right\}^{\frac{1}{2}}\left\{\dfrac{1}{n-k}\displaystyle\sum_{j=1}^{n-k} u_{j+k}^2 - \frac{1}{(n-k)^2}\left(\sum_{j=1}^{n-k} u_{j+k}\right)^2\right\}^{\frac{1}{2}}} \quad (30.7)$$

By convention we define

$$\left.\begin{aligned} r_0 &= 1 \\ r_{-k} &= r_k \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.8)$$

30.13. In practice we often require to calculate serial correlations up to r_{30} and for long series as many as 60. The arithmetic is tedious but may be systematised so as to reduce labour, which arises chiefly in the determination of cross-products forming the covariances.

The series of n terms is written down vertically on each of two slips of paper, the spacing being equal on the two slips. This can very conveniently be done on a Burroughs tabulator with a split keyboard, the series being recorded in duplicate and the resulting strip cut up

* It is sometimes convenient to confine this expression to values calculated from samples, the corresponding values for the infinite series being termed "autocorrelations" and denoted by a Greek ρ.

the middle. To calculate the first product-sum we pin the slips so that the first term on the right-hand slip is opposite the second term on the left-hand slip, and hence so that the jth term on the right is opposite to the $(j + 1)$th on the left all the way down. For most series the differences of two terms which are opposite can be obtained mentally by subtraction, squared, and set up on an adding-machine. The sum of squares of differences is thus determined, and the cross-product found from the simple identity

$$2 \Sigma (XY) = \Sigma (X^2) + \Sigma (Y^2) - \Sigma (X - Y)^2.$$

We then move the right-hand slip down one space so that the jth term is opposite the $(j + 2)$th term on the left and repeat the process; and so on to as many terms as may be required.

In this process $\Sigma (X^2)$ and $\Sigma (Y^2)$ are required at each stage, and it is as well to determine them by cumulative summation from the two ends of the series. $\Sigma (X)$ and $\Sigma (Y)$ are also required. It is also convenient on occasion to reduce the series to zero mean approximately before beginning the analysis.

Example 30.1

To illustrate the arithmetic we will take a very trivial example which the reader should check for himself. Take the series

$$-5, \ -6, \ -2, \ 4, \ 7, \ 3, \ 1, \ -5, \ -1, \ 2.$$

We set up the following scheme of tabulation for calculating serial correlations up to the fifth order :—

$n - k$.	k.	$\Sigma (X)$ (from beginning of series).	$\Sigma (Y)$ (from end of series).	$\Sigma (X^2)$ (from beginning).	$\Sigma (Y^2)$ (from end).	$\Sigma (X - Y)^2$.	$\Sigma (XY)$.
10	0	-2	-2	170	170	0	170
9	1	-4	3	166	145	143	84
8	2	-3	9	165	109	344	-35
7	3	2	11	140	105	445	-100
6	4	1	7	139	89	380	-76
5	5	-2	0	130	40	172	-1

The number $n - k$ is the number of pairs entering into the kth correlation. $\Sigma (X)$ is the sum of $n - k$ terms beginning at the first term, $\Sigma (Y)$ the corresponding sum of the last $n - k$ terms, and similarly for $\Sigma (X^2)$ and $\Sigma (Y^2)$. These are the quantities required to calculate the variances entering into the denominator of the kth serial correlation. The quantities $\Sigma (X - Y)^2$ are calculated by the moving-slip method described above.

We now calculate the correlation coefficients in the usual way, e.g. for r_1

$$\text{var } X = \frac{166}{9} - \left(-\frac{4}{9}\right)^2 = 18{\cdot}247$$

$$\text{var } Y = \frac{145}{9} - \left(\frac{3}{9}\right)^2 = 16{\cdot}000$$

$$\text{cov } (X, Y) = \frac{84}{9} - \left(-\frac{4}{9}\right)\left(\frac{3}{9}\right) = 9{\cdot}4815$$

$$r_1 = \frac{9{\cdot}4815}{\sqrt{(18{\cdot}247 \times 16)}} = +0{\cdot}55\,;$$

and for r_5

$$\text{var } X = \frac{130}{5} - \left(-\frac{2}{5}\right)^2 = 25{\cdot}840$$

$$\text{var } Y = \frac{40}{5} - \left(\frac{0}{5}\right)^2 = 8{\cdot}000$$

$$\text{cov }(X, Y) = -\frac{1}{5} - \left(-\frac{2}{5}\right)\left(\frac{0}{5}\right) = -\ 0{\cdot}200$$

$$r_5 = -\ 0{\cdot}01.$$

When n is large and the origin is chosen so that the mean of the whole series is approximately zero, a sufficiently good value of r is given by $\frac{\Sigma\,(XY)}{\{\Sigma\,(X^2)\,\Sigma\,(Y^2)\}^{\frac{1}{2}}}$, the corrections required to adjust the sums of squares and products to values about the mean being small; but this approximation must be used with some care and in any case the first two or three serial coefficients should be worked out exactly.

The Correlogram

30.14. The diagram obtained by graphing r_k as ordinate against k as abscissa and joining the points each to the next is called a *correlogram.* We shall give a number of examples below and shall see that the form of the correlogram provides a method of discriminating between the various types of oscillatory series.

30.15. Suppose, for example, that the series is generated by a moving average of random elements with weights $a_1, a_2, \ldots a_m$. The typical term of the series is then

$$u_j = a_1\,\varepsilon_j + a_2\,\varepsilon_{j+1} + \ldots + a_m\,\varepsilon_{j+m-1} \qquad . \quad . \quad . \ (30.9)$$

Without loss of generality we may take $E\,(\varepsilon) = 0$ and hence $E\,(u_j) = 0$. Then

$$E\,(u_j\,u_{j+k}) = E\,\{a_1\,\varepsilon_j + a_2\,\varepsilon_{j+1} + \ldots + a_m\,\varepsilon_{j+m-1}\}$$
$$\{a_1\,\varepsilon_{j+k} + a_2\,\varepsilon_{j+k+1} + \ldots + a_m\,\varepsilon_{j+k+m-1}\}.$$

Since

$$E\,(\varepsilon_j\,\varepsilon_{j+k}) = 0,\ k \neq 0$$
$$= v,\ \text{say, if } k = 0$$

we have

$$E\,(u_j\,u_{j+k}) = (a_1\,a_{k+1} + a_2\,a_{k+2} + \ldots + a_{m-k}\,a_m)\,v, \qquad . \quad . \ (30.10)$$

provided that $m > k$. But if $k \geqslant m$ then

$$E\,(u_j\,u_{j+k}) = 0. \qquad . \quad . \quad . \quad . \quad . \ (30.11)$$

Thus for an *infinite* series generated by the moving average the serial correlations vanish for $k \geqslant m$, and the correlogram from that point onwards coincides with the x-axis. In particular, if the a's are all equal to $1/m$, we have

$$E\,(u_j\,u_{j+k}) = (m - k)\,\frac{v}{m^2},$$

and hence

$$r_k = 1 - \frac{k}{m}, \qquad . \quad . \quad . \quad . \quad . \ (30.12)$$

so that the correlogram consists of a straight line joining the point (0, 1) to (k, 0), together with the x-axis from the latter point onwards.

Example 30.2

The weights of the Spencer 21-point formula are

$$\frac{1}{350}\{-1,\ -3,\ -5,\ -5,\ -2,\ 6,\ 18,\ 33,\ 47,\ 57,\ \mathbf{60},\ \ldots\}.$$

Apart from the divisor 350, which may be disregarded for present purposes, the sum of squares of weights is 17,542. The products (30.10) and the corresponding serial correlations are as follows :—

k.	$\Sigma\, a_j\, a_{j+k}$.	r_k.	k.	$\Sigma\, a_j\, a_{j+k}$.	r_k.
0	17,542	1·000	11	− 930	− 0·053
1	16,786	0·957	12	− 528	− 0·030
2	14,667	0·836	13	− 214	− 0·012
3	11,584	0·660	14	− 27	− 0·002
4	8,085	0·461	15	50	0·003
5	4,726	0·269	16	59	0·003
6	1,951	0·111	17	40	0·002
7	6	0·000	18	19	0·001
8	− 1,074	− 0·061	19	6	0·000
9	− 1,430	− 0·082	20	1	0·000
10	− 1,298	− 0·074	21	0	0·000

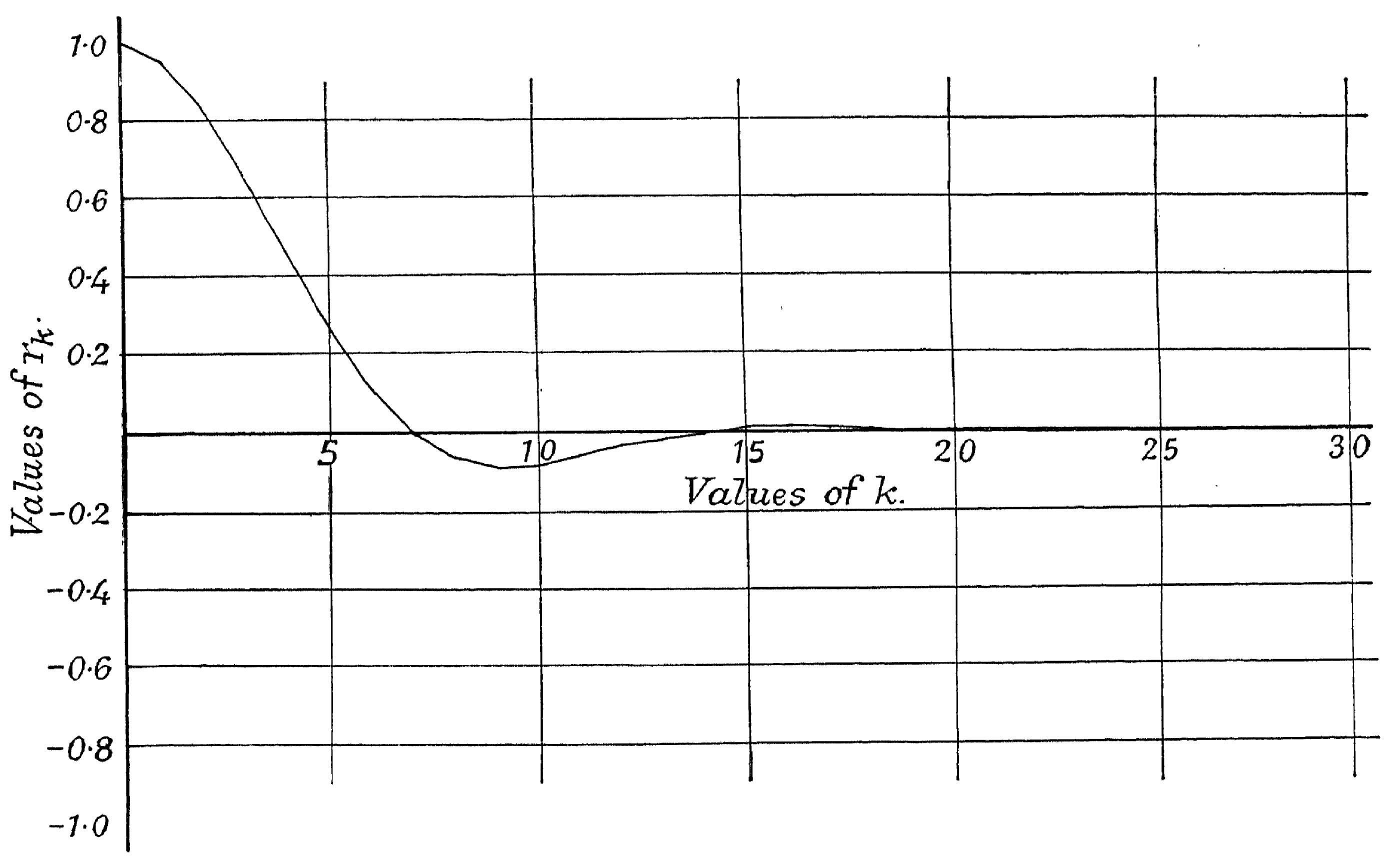


FIG. 30.3.—Correlogram of Series generated by the Spencer 21-point Formula (Example 30.2).

The correlogram is shown in Fig. 30.3. From $k = 13$ onwards the correlations are very small, and from $k = 21$ onwards they vanish completely.

30.16. Suppose now that the series consists of a sine term $A \sin \theta t$ plus ε_t, a random residual. As before, we may suppose $E(u_t) = 0$, and hence

$$\begin{aligned} E(u_j u_{j+k}) &= E\{A \sin \theta j + \varepsilon_j\}\{A \sin \theta (j+k) + \varepsilon_{j+k}\} \\ &= A^2 E\{\sin \theta j \sin \theta (j+k)\} \\ &= \frac{A^2}{n} \sum_{j=1}^{n} \{\sin \theta j \sin \theta (j+k)\} \qquad (30.13) \\ &= \frac{A^2}{2n} \Sigma \{\cos \theta k - \cos \theta (2j+k)\} \\ &= \frac{A^2}{2} \cos \theta k - \frac{A^2}{2n} \frac{\cos \theta (k+n+1) \sin n\theta}{\sin \theta}. \qquad (30.14) \end{aligned}$$

Thus for large n we have effectively, unless θ is small,

$$E(u_j u_{j+k}) = \frac{A^2}{2} \cos \theta k = B \cos \theta k, \text{ say.} \qquad (30.15)$$

Similarly we find

$$E(u_j^2) = B + \text{var } \varepsilon = C, \text{ say.} \qquad (30.16)$$

Hence

$$r_k = \frac{B}{C} \cos \theta k, \qquad k > 0. \qquad (30.17)$$

In short, for an infinite cyclical series the correlogram itself is a harmonic with period equal to that of the original harmonic component.

30.17. When the original series is the sum of several harmonic terms the formula for r_k will, in general, be the sum of harmonics, not necessarily with the same periods. Thus the correlogram will present a sinusoidal form which will not degenerate to the x-axis after some fixed point and will not, in fact, be damped.

30.18. Consider now the series defined by (30.3), namely

$$u_{t+2} = -au_{t+1} - bu_t + \varepsilon_{t+2}.$$

This is a difference equation which is easily solved by the usual methods.* The general solution of

$$u_{t+2} + au_{t+1} + bu_t = 0 \qquad (30.18)$$

is

$$u_t = p^t (A \cos \theta t + B \sin \theta t) \qquad (30.19)$$

where

$$\left.\begin{aligned} p &= \sqrt{b} \\ \cos \theta &= -\frac{a}{2\sqrt{b}} \end{aligned}\right\} \qquad (30.20)$$

Here $\sqrt{b}$ is to be taken with positive sign, and it is assumed that $4b > a^2$. We also assume that $\sqrt{b}$ is not greater than unity. The contrary case is mathematically permissible, but it implies that u_t increases without limit, which is outside the domain of our consideration.

* See, for instance, Milne-Thomson, *Calculus of Finite Differences*, chapter 13.

Consider now the series

$$\sum_{j=0}^{\infty} \xi_j \, \varepsilon_{t-j+1}, \qquad . \quad . \quad . \quad . \quad . \quad (30.21)$$

where ξ_t is a particular solution of (30.19) such that $\xi_0 = 0$ and $\xi_1 = 1$, i.e. such that

$$\xi_t = \frac{2}{\sqrt{(4b - a^2)}} p^t \sin \theta t. \qquad . \quad . \quad . \quad . \quad (30.22)$$

On substituting (30.21) in the original equation it will be found to provide a particular solution. The general solution is then

$$u_t = p^t (A \cos \theta t + B \sin \theta t) + \sum_{j=0}^{\infty} \xi_j \, \varepsilon_{t-j+1}. \qquad . \quad . \quad . \quad (30.23)$$

As p is not greater than unity we shall, in general, find that the first term in this expression is damped out of existence. If we may regard our series as having been "started up" some time prior to the point $t = 0$, the solution is effectively

$$u_t = \sum_{j=0}^{\infty} \xi_j \, \varepsilon_{t-j+1}. \qquad . \quad . \quad . \quad . \quad . \quad (30.24)$$

30.19. In this form the autoregressive scheme is seen to be a moving average of a component ε with infinite extent and damped harmonic weights. Consider now its correlogram. We have

$$\sum_{j=0}^{\infty} \xi_j \, \xi_{j+k} = \frac{4}{4b - a^2} \Sigma \{ p^{2j+k} \sin \theta j \sin \theta (j + k) \}$$

$$= \frac{2p^k}{4b - a^2} \Sigma [p^{2j} \{ \cos \theta k - \cos \theta (2j + k) \}]$$

$$= \frac{2p^k}{4b - a^2} \left\{ \frac{\cos \theta k}{1 - p^2} - \frac{\cos \theta k - p^2 \cos \theta (k - 2)}{1 - 2p^2 \cos 2\theta + p^4} \right\}. \qquad . \quad (30.25)$$

Now

$$E (u_j \, u_{j+k}) = E \{ \sum_j (\xi_j \, \varepsilon_{t-j+1}) \sum_j (\xi_j \, \varepsilon_{t+k-j+1}) \}$$

$$= \text{var } \varepsilon \sum_{j=0}^{\infty} (\xi_j \, \xi_{j+k}) . \qquad . \quad . \quad . \quad . \quad . \quad (30.26)$$

Thus

$$r_k = \frac{\text{var } \varepsilon \sum_{j=0}^{\infty} (\xi_j \, \xi_{j+k})}{\text{var } \varepsilon \sum_{j=0}^{\infty} \xi_j^2},$$

which, on substitution from (30.25), reduces to

$$r_k = \frac{p^k}{(1+p^2)\sin\theta}\{\sin(k+1)\theta - p^2\sin(k-1)\theta\}. \qquad (30.27)$$

Writing

$$\tan\psi = \frac{1+p^2}{1-p^2}\tan\theta, \qquad (30.28)$$

we find

$$r_k = \frac{p^k\sin(k\theta+\psi)}{\sin\psi}, \qquad k \geqslant 0 . \qquad (30.29)$$

From this we see that the correlogram will oscillate with period $2\pi/\theta$, but that, owing to the factor p^k, *it will be damped.* If k is negative the formula applies, except that $|k|$ must be used instead of k on the right-hand side of (30.29).

30.20. We thus reach the interesting conclusion that the three types of series considered in **30.9**, however similar to the eye, will have distinct types of correlogram, provided that the series are long enough for the observed correlations to approach the expected values for an infinite series. The correlogram of a series generated by moving averages, though it may oscillate as in Example 30.2, will vanish after a certain point; that of a series of harmonic terms will oscillate, but will not vanish or be damped; that of the autoregressive scheme will oscillate and will not vanish, but it will be damped. The correlogram therefore offers a theoretical basis for discriminating between the three types of oscillatory series.

30.21. Unfortunately the series with which we have to work are very frequently too short to enable a decisive distinction to be made. We shall see below that divergence between theory and observation can be very considerable, and that sampling theory has not yet advanced far enough to enable us to make objective judgments in probability about its significance. We shall have to rely on limited experimental evidence and to some extent on intuitive judgment in reaching conclusions. If, therefore, the remainder of this chapter contains gaps in the treatment and leaves certain points undecided the reader will understand that the reason is ignorance rather than indifference.

Examples of Correlograms from Observed Series

30.22. We will in the first place give the correlograms of a few of the series given earlier in this and the preceding chapter.

Example 30.3

In Table 30.2 we gave the deviations from the trend of marriage rates for the years 1843–1896. The first 20 serial correlations of this series are shown in Table 30.5 and the correlogram in Fig. 30.4.

TABLE 30.5

Serial Correlations of the Marriage Data of Table 30.2.

Order of Correlation k.	r_k.	Order of Correlation k.	r_k.
1	0·563	11	− 0·080
2	− 0·089	12	− 0·136
3	− 0·498	13	− 0·132
4	− 0·631	14	− 0·058
5	− 0·467	15	− 0·095
6	− 0·025	16	− 0·126
7	0·353	17	− 0·036
8	0·396	18	0·131
9	0·254	19	0·209
10	0·104	20	0·205

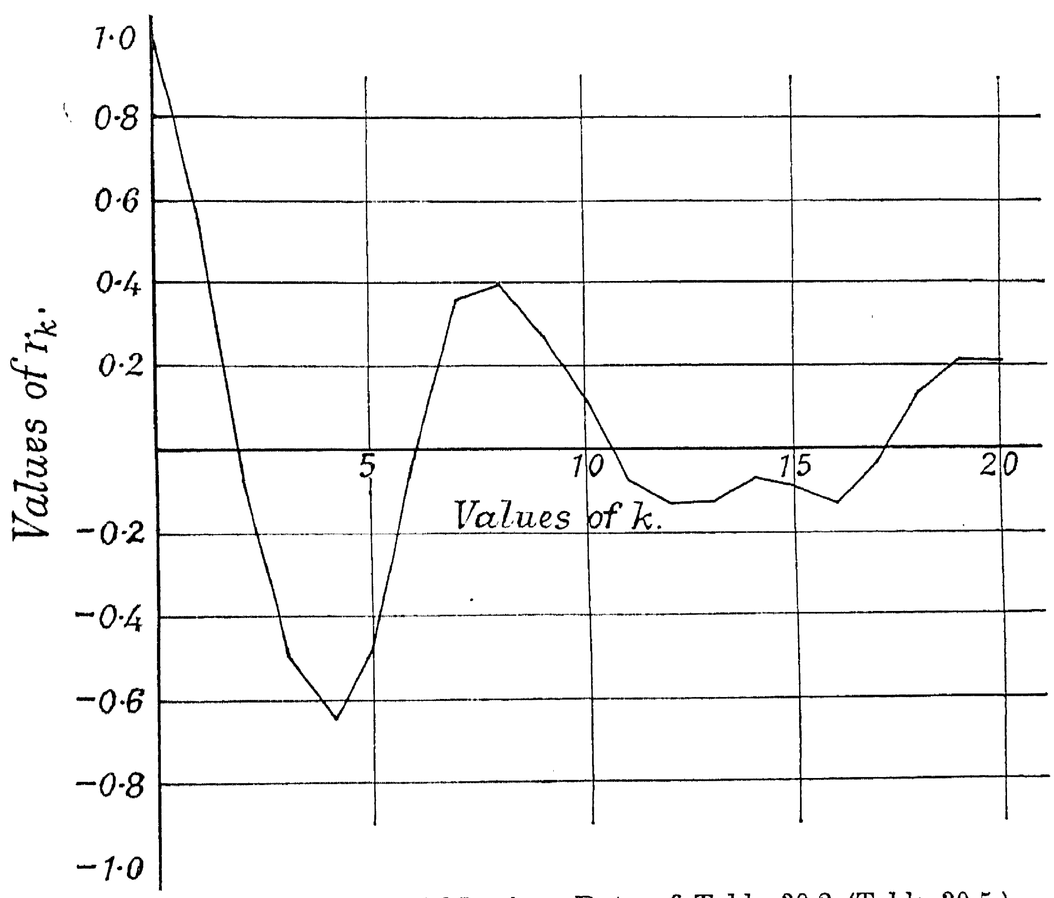


FIG. 30.4.—Correlogram of Marriage Data of Table 30.2 (Table 30.5.).

The correlogram is smooth and suggests the operation of an autoregressive scheme. There is little indication that a moving average, at least of extent less than 20, would account for the series, but on the other hand some damping appears to be present.

Example 30.4

Table 30.6 shows the first 60 serial correlations of the Beveridge series of Table 30.1, the correlogram being given in Fig. 30.5.

TABLE 30.6

Serial Correlations of the Beveridge Wheat-Price Index of Table 30.1.

Order of Correlation k.	r_k.	k.	r_k.	k.	r_k.	k.	r_k.
1	0·562	16	0·158	31	0·060	46	− 0·036
2	0·103	17	0·109	32	− 0·008	47	− 0·013
3	− 0·075	18	0·002	33	− 0·039	48	0·042
4	− 0·092	19	− 0·075	34	0·007	49	0·062
5	− 0·082	20	− 0·062	35	0·056	50	0·065
6	− 0·136	21	− 0·021	36	0·010	51	0·050
7	− 0·211	22	− 0·062	37	− 0·004	52	0·009
8	− 0·261	23	− 0·088	38	− 0·015	53	− 0·027
9	− 0·192	24	− 0·084	39	− 0·047	54	− 0·053
10	− 0·070	25	− 0·076	40	− 0·047	55	− 0·073
11	− 0·003	26	− 0·091	41	0·008	56	− 0·106
12	− 0·015	27	− 0·052	42	0·034	57	− 0·084
13	− 0·012	28	− 0·032	43	0·065	58	− 0·019
14	0·047	29	− 0·012	44	0·099	59	0·003
15	0·101	30	0·059	45	0·009	60	0·010

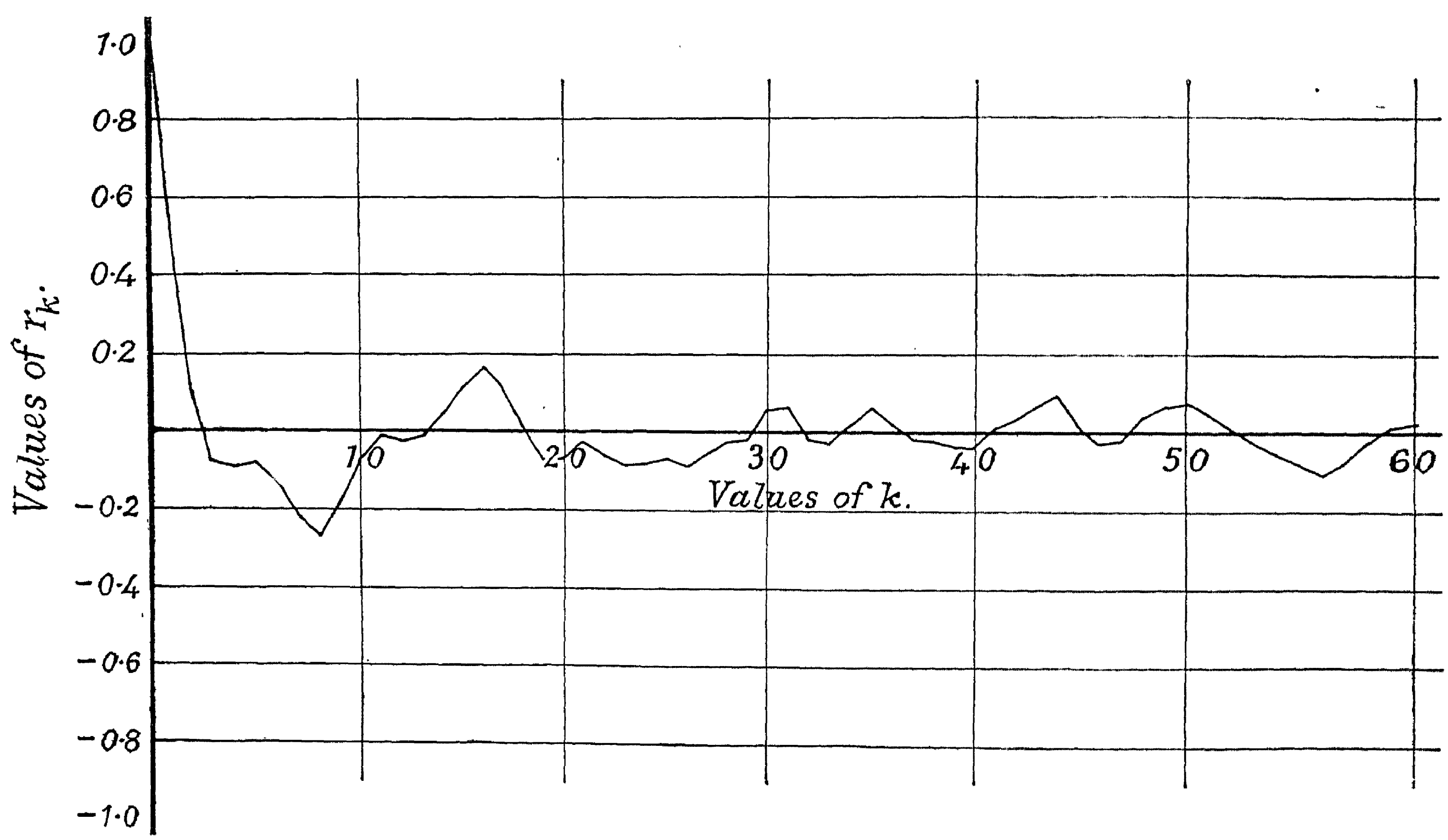


FIG. 30.5.—Correlogram of the Beveridge Series of Table 30.1 (Table 30.6).

The correlogram here is almost certainly damped. The oscillations persist in a most remarkable way, notwithstanding the diminishing amplitude, and the presumption is a strong one that the series is of the damped type.

Example 30.5

In Table 29.8 (page 386) we gave the residuals of a sheep-population series for the years 1871 to 1935. Table 30.7 shows the first 30 serial correlations of this series and Fig. 30.6 the correlogram. Again the correlogram is oscillatory, but the damping is not so clear.

TABLE 30.7

Serial Correlations of the Sheep Data of Table 29.8.

Order of Correlation k.	r_k.	k.	r_k.	k.	r_k.
1	0·595	11	− 0·142	21	− 0·381
2	− 0·151	12	− 0·172	22	− 0·118
3	− 0·601	13	− 0·186	23	0·173
4	− 0·537	14	− 0·128	24	0·343
5	− 0·138	15	0·052	25	0·352
6	0·144	16	0·276	26	0·154
7	0·203	17	0·439	27	− 0·203
8	0·118	18	0·293	28	− 0·456
9	0·006	19	− 0·074	29	− 0·415
10	− 0·078	20	− 0·359	30	− 0·184

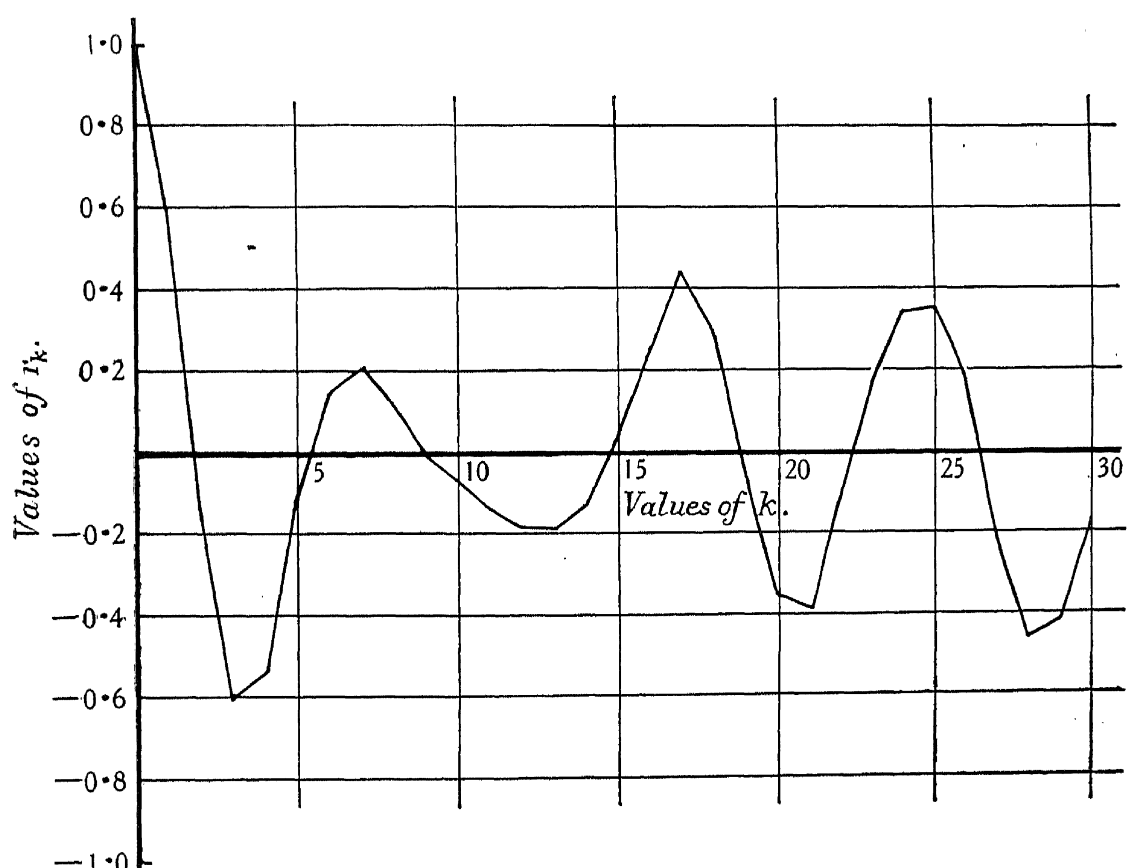


Fig. 30.6.—Correlogram of the Sheep Population Data of Table 29.8 (Table 30.7.)

Significance of a Correlogram

30.23. The foregoing examples illustrate one of the main difficulties we have to face in correlogram analysis. On intuitive grounds we seem to be justified in rejecting the scheme of moving averages as a possible scheme for the series of these examples, since the oscillations in the correlograms persist; but we can no doubt find moving averages which will produce such correlograms, though their extents would have to be long (over 60 in the case of the Beveridge series) and their weights artificial. The only final test seems to be to ascertain such a moving average and then to examine whether it will predict further terms in the series if such can be observed.

30.24. Distinction between the scheme of harmonic components and the autoregressive scheme is even more difficult for short series, since the correlograms for the latter do not damp out according to expectation. Consider in fact an autoregressive scheme of the simple linear type (30.3). There will be the usual variation in length from peak to peak and in amplitude; but if the section of the series is a comparatively short one, covering, say, four or five oscillations, the oscillations will not have time to get very much out of step and the serial correlations will be systematically larger than one would expect for an infinite series. This effect is exhibited in Table 30.8 and Fig. 30.7, which give the serial correlations and the correlogram for the series of Table 30.4, given by the formula

$$u_{t+2} = 1{\cdot}1\ u_{t+1} - 0{\cdot}5\ u_t + \varepsilon_{t+2}.$$

Here the damping factor $p = \sqrt{b} = 0{\cdot}7071$, and by the thirtieth correlation r_k should be very small, less than $0{\cdot}002$ in absolute magnitude. Actually it is 100 times as large. The mere fact that an observed correlogram for a short series fails to damp very rapidly is not, therefore, a very definite indication that the series is not ruled by the autoregressive scheme. On the contrary, failure to damp may be expected.

30.25. We are on firmer ground when considering the significance of a correlogram in the sense of judging whether it can be derived from a random series.

(*a*) The variance of r_k in a random series of n terms is approximately $\dfrac{1}{n-k}$, provided that n is large. For

$$E\left\{\frac{1}{n-k}\sum_{j=1}^{n-k}(x_j\,x_{j+k})\right\}^2 = \frac{1}{(n-k)^2}E\ \{\Sigma\, x_j^2\, x_{j+k}^2 + 2\Sigma\, x_j\, x_{j+k}\, x_m\, x_{m+k}\}, \qquad j \neq m$$

$$= \frac{1}{(n-k)^2}E\ \Sigma\ (x_j^2\, x_{j+k}^2)$$

$$= \frac{1}{n-k}\operatorname{var}^2 x.$$

Hence, for large samples,

$$\operatorname{var} r = \frac{1}{n-k}\,\frac{\operatorname{var}^2 x}{\operatorname{var}^2 x} = \frac{1}{n-k}. \qquad (30.30)$$

R. L. Anderson (1942) has recently given exact results for the significance of a serial correlation.

(*b*) For our purposes, however, the important point is not whether a particular serial coefficient is significant, but whether the oscillatory character of the correlogram as a whole

TABLE 30.8

Serial Correlations of the Artificial Series of Table 30.4.

Order of Correlation k.	r_k.	k.	r_k.	k.	r_k.
1	0·70	11	− 0·05	21	0·05
2	0·29	12	− 0·17	22	− 0·12
3	0·01	13	− 0·27	23	− 0·28
4	− 0·17	14	− 0·31	24	− 0·43
5	− 0·27	15	− 0·30	25	− 0·57
6	− 0·25	16	− 0·18	26	− 0·56
7	− 0·13	17	0·12	27	− 0·26
8	0·07	18	0·29	28	0·02
9	0·12	19	0·33	29	0·17
10	0·05	20	0·22	30	0·27

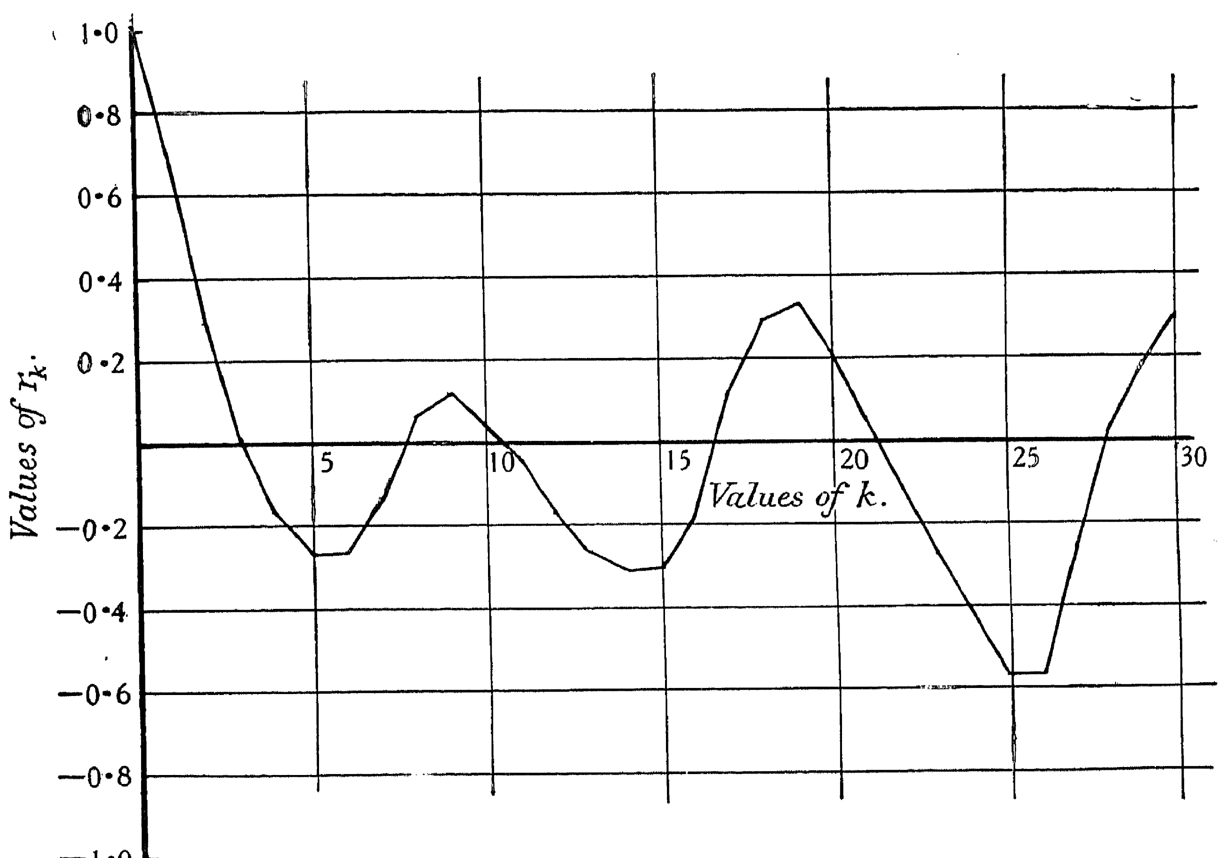


FIG. 30.7.—Correlogram of the Artificial Series of Table 30.4 (Table 30.8.).

is so. Here we have to form an intuitive judgment, but it can hardly be doubted that the undulations in Figs. 30.4 to 30.6 are not accidental. Something exists to be explained as a systematic effect, though what that effect is may be more difficult to decide.

30.26. We shall proceed to study the autoregressive scheme and the scheme of cyclical components in more detail, without prejudice for the time being to the question as to which is the better representation in particular cases. This latter is not, in fact, entirely a statistical matter, and we shall return to it in **30.39**.

The Autoregressive Scheme

30.27. We consider in the first instance the simplified scheme of equation (30.3). The theoretical correlogram for a series generated by this equation is of the damped type given by (30.29),

$$r_k \doteq \frac{p^k \sin (k\theta + \psi)}{\sin \psi},$$

where $2\pi/\theta$ is the *autoregressive period* of the regression equation and is given by

$$\cos \theta = -\frac{a}{2\sqrt{b}}.$$

The typical series of this kind has no "period" in the strict sense. The lengths from peak to peak or from upcross to upcross vary in the characteristic way. It appears from experiment (but has not, I think, been shown theoretically) that the distribution of distances from peak to peak is of the unimodal type with a central value somewhere near the mean distance between peaks; and similarly for troughs and upcrosses. In speaking of the "period" of an autoregressive series we mean the central value of one of these distributions. The question we have now to consider is whether this period is the same as the autoregressive period $2\pi/\theta$ of the regression equation.

30.28. We have seen in **29.26** that the mean distance between upcrosses of the series generated by the moving average whose weights are $\xi_1 \ldots \xi_m$ is given by $2\pi/\phi$, say, where

$$\cos \phi = \frac{\sum_{j=1}^{m-1} \xi_j \xi_{j+1}}{\sum_{j=1}^{m} \xi_j^2}.$$

Substituting for ξ from (30.22) and using (30.25), we find

$$\begin{aligned}\cos \phi &= \frac{\dfrac{2p}{4b - a^2}\left\{\dfrac{\cos \theta}{1 - p^2} - \dfrac{\cos \theta\,(1 - p^2)}{1 - 2p^2 \cos 2\theta + p^4}\right\}}{\dfrac{2}{4b - a^2}\left\{\dfrac{1}{1 - p^2} - \dfrac{1 - p^2 \cos 2\theta}{1 - 2p^2 \cos 2\theta + p^4}\right\}} \\ &= \frac{2p \cos \theta}{1 + p^2} \\ &= -\frac{a}{1 + b}. \qquad (30.31)\end{aligned}$$

Thus the mean period as defined by upcrosses is

$$2\pi/\text{arc cos}\left(\frac{-a}{1 + b}\right) \qquad (30.32)$$

whereas that for the autoregressive period of the equation is

$$2\pi/\text{arc cos}\left(\frac{-a}{2\sqrt{b}}\right). \qquad (30.33)$$

30.29. The mean period between upcrosses is thus *not* the same as the autoregressive period. The two are very close for many of the values of a and b arising in practice. For instance, when $b = 1$ they are identical; when $a = 1$, $b = 0{\cdot}5$ their ratio is 1·07. One might infer that an estimate of the period of an autoregressive scheme can be obtained from the correlogram, but this generalisation requires some important qualifications.

(*a*) Firstly, the ratio of (30.33) to (30.32) is not necessarily close to unity for values of b in the neighbourhood of $a^2/4$, i.e. when θ is small and the autoregressive period is long. Consider, for instance, the series generated by

$$u_{t+2} = 1{\cdot}2u_{t+1} - 0{\cdot}4u_t + \varepsilon_{t+2}.$$

We have

$$\cos\theta = -\frac{a}{2\sqrt{b}} = \frac{1{\cdot}2}{2\sqrt{0{\cdot}4}} = 0{\cdot}9499$$

$$\theta = 18{\cdot}2^\circ, \qquad \text{period} = 19{\cdot}7 \text{ units.}$$

However, for ϕ,

$$\cos\phi = \frac{1{\cdot}2}{1{\cdot}4} = 0{\cdot}8571$$

$$\phi = 31^\circ, \qquad \text{period} = 11{\cdot}6 \text{ units.}$$

The mean distance between upcrosses, and *a fortiori* that between peaks, is very much shorter than the autoregressive period.

(*b*) The mean distance between upcrosses may miss certain oscillations above or below the x-axis, so that it overestimates the period between peaks or troughs. On the other hand, the latter may include ripples on the main wave which we wish to ignore. The reader can verify for himself, by constructing an autoregressive series by some such formula as the above, how difficult it is to draw the line in particular cases. The difficulty, however, must be faced, for it is precisely the kind which we meet in dealing with observed series.

(*c*) Owing to the appearance of the phase angle ψ in equation (30.29) the starting-point of the correlogram ($k = 0$) is not to be regarded as a maximum. The period of the correlogram is therefore to be calculated either by ignoring this point or by reference to distances between troughs and upcrosses in the correlogram.

30.30. The equation

$$u_{t+2} + au_{t+1} + bu_t = \varepsilon_{t+2}$$

may be regarded as expressing the regression of u_{t+2} on u_{t+1} and u_t, the term ε_{t+2} being a residual error. We may therefore estimate the constants a and b from the regression equation of the observed series in the usual way. If we assume that the series is long enough for end effects to be negligible in determining the variances of the finite series, then var u_{t+2} = var u_{t+1} = var u_t, and from the usual formulae for regressions we find

$$a = -\frac{r_1(1 - r_2)}{1 - r_1^2} \qquad \qquad (30.34)$$

$$b = -\frac{r_2 - r_1^2}{1 - r_1^2} = -1 + \frac{1 - r_2}{1 - r_1^2}. \qquad \qquad (30.35)$$

This gives us the constants of the autoregressive scheme from the serial correlations.

It should, however, be realised that these estimates are rather sensitive to superposed error of the type we refer to below (**30.32**), and it is therefore unsafe to estimate the

autoregressive period from them. The correlogram itself appears to be a safer guide on this matter.

Example 30.6

Consider again the sheep data of Table 30.7 and Fig. 30.6. Suppose we have decided, from the appearance of the correlogram, to attempt to represent the series by an autoregressive scheme.

In the first place, we have to inquire whether a scheme of the simple linear form (30.3) is likely to be adequate. Would it, for example, be better to consider the more general form

$$u_{t+3} + au_{t+2} + bu_{t+1} + cu_t = \varepsilon_{t+3},$$

or need we take into account curvilinear regressions such as

$$u_{t+2} + au_{t+1} + a' u_{t+1}^2 + bu_t + b' u_t^2 + \varepsilon_{t+2} \; ?$$

The first point can be elucidated by the use of partial and multiple correlations. The following are the partial coefficients and the function of the multiple correlation $1 - R^2$ as determined by the continued product of $(1 - r^2)$ (cf. vol. I, equation 15.45, p. 380) :—

Order of Partial Correlation.	Value of Partial Correlation.	$\Pi (1 - r^2)$.
12	0·595	0·6460
13.2	− 0·782	0·2509
14.23	0·097	0·2485
15.234	− 0·183	0·2402
16.2345	0·031	0·2400
17.23456	0·014	0·2400

Evidently no appreciable gain in representation is to be obtained by taking the regression on more than the two preceding terms.

The possibility as to better representation by taking curvilinear regressions may be considered by drawing the scatter diagrams of u_t on u_{t+1} and u_t on u_{t+2}. These are shown in Fig. 30.8. It seems clear that there is an essential scatter in the data which no ordinary polynomial can represent, and that curvilinear terms are unlikely to add anything material to the linear regressions.

We conclude that if the data are of the autoregressive type it is unnecessary to consider any more elaborate scheme than the simple type

$$u_{t+2} + au_{t+1} + bu_t = \varepsilon_{t+2}.$$

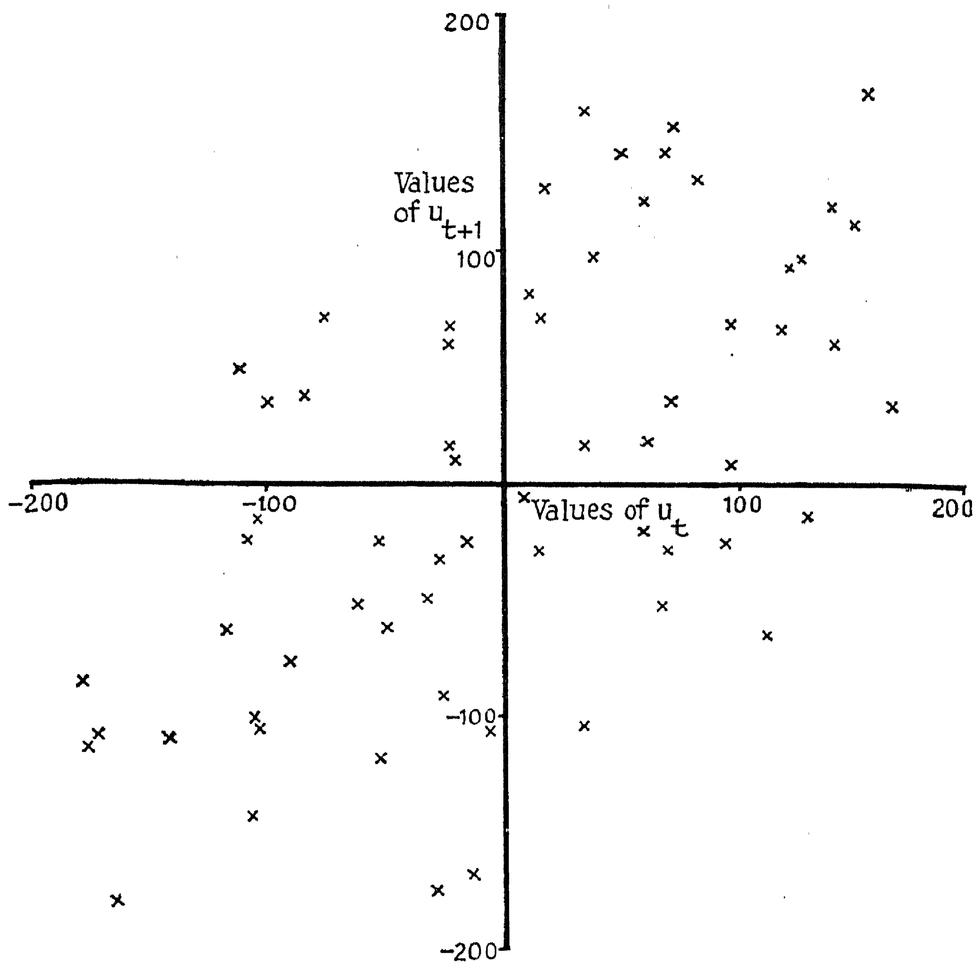
For this series we have

$$r_1 = 0{\cdot}595, \qquad r_2 = -\,0{\cdot}151.$$

Hence

$$-\,a = \frac{r_1 (1 - r_2)}{1 - r_1^2} = 1{\cdot}060$$

$$-\,b = \frac{r_2 - 1}{1 - r_1^2} + 1 = -\,0{\cdot}782.$$



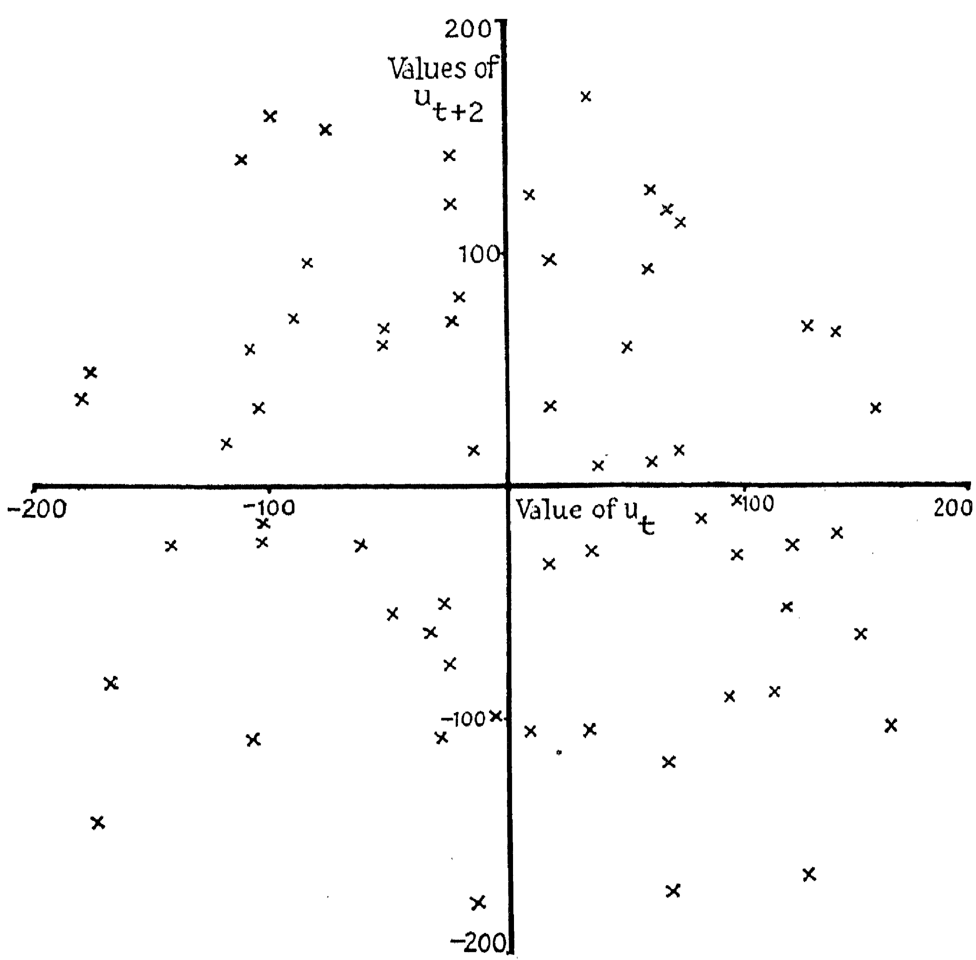


FIG. 30.8.—Scatter Diagrams of u_t on u_{t+1} (top figure), and u_t on u_{t+2} (bottom figure).

The autoregression equation is

$$u_{t+2} = 1{\cdot}060\ u_{t+1} - 0{\cdot}782\ u_t + \varepsilon_{t+2}.$$

For the autoregressive period we have

$$\cos\theta = \frac{1{\cdot}060}{2\sqrt{(0{\cdot}782)}} = 0{\cdot}600, \qquad \theta = 53{\cdot}2^\circ$$

and hence the period is $\dfrac{360}{53{\cdot}2} = 6{\cdot}8$ years.

Now in the correlogram (Fig. 30.6) there are peaks at $k = 7$, 17 and 25, giving a period of about 9 years; and there are troughs at $k = 3$, 13, 21 and 28, giving a mean period of 8·3 years. The autoregressive period as estimated from the correlogram is then between 8 and 9 years, whereas that given by the autoregression equation is 6·8 years, considerably shorter.

Using the values of a and b found above, we have for the mean distance between upcrosses,

$$\cos\phi = \frac{1{\cdot}060}{1{\cdot}782} = 0{\cdot}5948, \qquad \phi = 53{\cdot}5^\circ,$$

giving a mean distance practically equal to the autoregressive period as shown by the regression equation.

Finally, looking to the original series, we see that there are nine major peaks, the first in 1874 and the last in 1932, so that the mean distance between peaks is $\dfrac{58}{8} = 7{\cdot}25$ years; and nine upcrosses, the first between 1872 and 1873 and the last between 1930 and 1931, so that the mean distance between upcrosses is $\dfrac{58}{8} = 7{\cdot}25$ years, the same as for peaks. The upcross at 1876–7, however, is due to a temporary fall below the zero line, and had it not occurred we should have found a mean distance of 8·3 years.

We have therefore reached this position: the mean period in the series itself appears to be about 7·25 years; that given by the regression constants is 6·8 years; and that given by the correlogram is about 8·5 years. These figures are scarcely close enough for comfort, and further data would be required to arrive at a more accurate estimate of the mean period. Nevertheless, they illustrate very well the kind of divergence which appears to be more the rule than the exception in dealing with short series. We should expect the correlogram to give a higher value than the series itself, for there may appear peaks or upcrosses in the latter which are purely temporary fluctuations due to the casual element. On the other hand, the regression constants appear to give consistently lower values for the autoregressive period than the correlogram, an effect found by Yule (1927*a*) for sunspots, Wold (1938*a*) for cost-of-living indices, and Kendall (1944*a*) in series of agricultural prices, acreage and livestock populations.

30.31. Let us examine more closely the effect referred to at the end of the previous example. Our autoregressive system is based on a random element ε_t which is added to the term u_{t+2}. We can therefore regard the value at time $t + 2$ as composed of two parts, a systematic element expressed by $au_{t+1} + bu_t$, giving the effect of the past history of the system at times $t + 1$ and t, together with a new random element peculiar to the moment. This latter is random in the sense that it is casual and unpredictable; but once it has occurred it is incorporated into the motion of the system and exerts an influence on future

history. It is therefore quite unlike an error of observation or a sampling error which distorts the value of a particular member but does not affect the others.

Now suppose that such an error of observation is present, and let us represent it by η. For long series this element will increase the variance of the observed values by var η, but if it is independent of the remaining constituents of the series it will not affect the covariances. Hence the serial correlations will all be reduced in a constant proportion c, except of course r_0; and this, as we proceed to show, will affect the autoregressive period as derived from the regression constants, in general shortening the period quite considerably.

30.32. If r_1 is reduced to cr_1 and r_2 to cr_2, the constants of the regression equations are, from (30.34) and (30.35),

$$-a' = \frac{cr_1(1-cr_2)}{1-c^2r_1^2} \qquad (30.36)$$

$$-b' = \frac{cr_2 - c^2r_1^2}{1-c^2r_1^2}. \qquad (30.37)$$

The estimated autoregressive period is then θ', given by

$$\cos\theta' = -\frac{a'}{2\sqrt{b'}}$$

$$= \frac{cr_1(1-cr_2)}{2\sqrt{(1-c^2r_1^2)(c^2r_1^2-cr_2)}}.$$

Differentiating the logarithm of this expression and putting $c = 1$, we find

$$-2\tan\theta'\frac{d\theta'}{dc} = 1 - \frac{2r_2}{1-r_2} + \frac{2r_1^2}{1-r_1^2} + \frac{r_1^2}{r_2-r_1^2},$$

which reduces to

$$-\tan\theta'\frac{d\theta'}{dc} = \frac{(1+b)(3b^2+b-a^2)}{2b\{(1+b)^2-a^2\}}. \qquad (30.38)$$

Now $\tan\theta = \sqrt{\left(\frac{4b}{a^2}-1\right)}$ and the period $P = 2\pi/\theta$. We then find

$$\left(\frac{dP}{dc}\right)_{c=1} = \frac{P^2a(1+b)(3b^2+b-a^2)}{4\pi b\{(1+b)^2-a^2\}\sqrt{(4b-a^2)}}. \qquad (30.39)$$

This equation gives us an approximate idea of the change in the period P for small changes in c near $c = 1$. For instance, with $a = -1{\cdot}5$, $b = 0{\cdot}9$ we find $P = 9{\cdot}7$ units, and from (30.39),

$$\left(\frac{dP}{dc}\right)_{c=1} = -16{\cdot}5.$$

Thus, if $c = 0{\cdot}9$, i.e. the variance of η is about 10 per cent. of the total, the period will be reduced by about 1·65 years, a substantial amount.

30.33. It is thus possible that the observed discrepancies between the autoregressive periods as given by the regression constants and the correlogram may be due to *superposed* random fluctuation which is not incorporated into the autoregressive scheme. This is not the only possible explanation; for instance, in particular cases the disturbance function ε may not be random. The hypotheses to be considered in such a case, however, are so complex that it is difficult to pursue a quantitative investigation without a wealth of material; and this, unfortunately, is usually denied to us, at least in economic work.

Meteorological data are more numerous, and we may hope that further light will be thrown on the autoregressive scheme by a re-examination of the material available in this field.

30.34. Consider now the more extended autoregression equation

$$u_{t+m} + a_1 u_{t+m-1} + a_2 u_{t+m-2} + \ldots + a_m u_t = \varepsilon_{t+m}. \qquad (30.40)$$

The explicit solution cannot be given in the simple form available when $m = 2$. It has, in general, the solution

$$u_t = A_1 \alpha_1^t + A_2 \alpha_2^t + \ldots + A_m \alpha_m^t + B, \qquad (30.41)$$

where $\alpha_1 \ldots \alpha_m$ are the roots of

$$\alpha^m + a_1 \alpha^{m-1} + a_2 \alpha^{m-2} + \ldots + a_m = 0, \qquad (30.42)$$

and B is a particular integral involving the ε's. For the series to be oscillatory without increasing indefinitely no term such as x^t, where x is real and greater than unity, can appear. Assuming this to be so, and assuming further that the series was "started up" some time before $t = 0$, we reduce the solution to the particular integral B.

Choose a particular value ξ_t of $\sum_{j=1}^{m} A_j \alpha_j^t$, such that

$$\left.\begin{aligned} \xi_0 &= 0 \\ \xi_1 + a_1 \xi_0 &= 1 \\ \xi_2 + a_1 \xi_1 + a_2 \xi_0 &= 0 \\ \cdot \quad \cdot \quad &\cdot \quad \cdot \\ \xi_{m-1} + a_1 \xi_{m-2} + \ldots + a_{m-1} \xi_0 &= 0. \end{aligned}\right\} \qquad (30.43)$$

This is always possible in general, for it imposes m conditions on the m constants A. Then it will be found on substitution that a particular integral B is given by

$$u_t = \sum_{j=0}^{\infty} \xi_j \varepsilon_{t-j+1}, \qquad (30.44)$$

a generalisation of (30.24). Our series may then be regarded as generated by a moving average of infinite extent, the weights being combinations of damped harmonic and exponential terms.

30.35. The correlogram of such a series may be determined by the following method, due to Walker (1931). Multiply (30.40) by u_{t-k} and sum. We find

$$r_{k+m} + a_1 r_{k+m-1} + a_2 r_{k+m-2} + \ldots + a_m r_k = \frac{\Sigma (\varepsilon_{t+m} u_{t-k})}{\text{var } u}. \qquad (30.45)$$

Now u_{t-k} depends only on ε_{t-k} and terms with lower subscripts and hence is uncorrelated with ε_{t+m} for $k > -m$. Thus we have

$$r_{k+m} + a_1 r_{k+m-1} + \ldots + a_m r_k = 0, \qquad k > -m. \qquad (30.46)$$

If we multiply (30.40) by u_{t+k+m} we find similarly

$$r_k + a_1 r_{k+1} + \ldots + a_m r_{k+m} = \frac{\Sigma (\varepsilon_{t+m} u_{t+k+m})}{\text{var } u} \qquad (30.47)$$

but the expression on the right no longer vanishes. In fact u_{t+k+m} contains the term $\xi_{k+1} \varepsilon_{t+m}$, and hence

$$r_k + a_1 r_{k+1} + \ldots + a_m r_{k+m} = \xi_{k+1} \frac{\text{var } \varepsilon}{\text{var } u}, \qquad k \geqslant -m. \qquad (30.48)$$

From (30.46) it follows that the serial correlation r_k will be given by

$$r_k = \sum_j (A_j \alpha_j^k), \qquad (30.49)$$

where the α's are the roots of (30.42) and the A's are constants to be determined from initial conditions. Thus the correlogram will be the sum of terms which either decay exponentially to zero (α real) or oscillate with a similar decay to zero (α complex). Walker (1931) has used this result in an inquiry into a series of atmospheric pressures.

The Autocorrelation Function

30.36. If we have a series $u(t)$ defined at every point of time in some range $-h$ to $+h$, we may define its variance as

$$\frac{1}{2h}\int_{-h}^{h} u^2(t)\, dt \qquad (30.50)$$

on the assumption that the mean value is zero, which does not limit our generality. Suppose the series is reduced to standard measure by dividing throughout by the square root of this variance. Then an evident generalisation of the serial correlation is given by

$$r(k) = \frac{1}{2h}\int_{-h}^{h} u(t)\, u(t+k)\, dt. \qquad (30.51)$$

We shall call this the *autocorrelation function.* We can likewise regard it as defined when h tends to infinity, provided that the limit on the right in (30.51) exists. It is to be noted that $r(k)$ is in that case an even function of k.

30.37. We shall also consider the function

$$R(k) = \int_{-\infty}^{\infty} u(t)\, u(t+k)\, dt, \qquad (30.52)$$

when it exists. We have

$$\int_{-\infty}^{\infty} R(k)\, e^{ikp}\, dk = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ikp}\, u(t)\, u(t+k)\, dt\, dk$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ip(t+k)}\, u(t+k)\, e^{-ipt}\, u(t)\, dt\, dk.$$

The simple substitution $t + k = q$ reduces this to

$$\int_{-\infty}^{\infty} e^{ipq}\, u(q)\, dq \int_{-\infty}^{\infty} e^{-ipt}\, u(t)\, dt.$$

Thus, if we write

$$\alpha(p) + i\beta(p) = \int_{-\infty}^{\infty} e^{ipq}\, u(q)\, dq, \qquad (30.53)$$

we have

$$\int_{-\infty}^{\infty} R(k)\, e^{ikp}\, dk = \alpha^2(p) + \beta^2(p). \qquad (30.54)$$

It follows, as is otherwise evident from the fact that $R(k)$ is an even function, that the imaginary part on the left of (30.54) vanishes, and we have

$$\int_{-\infty}^{\infty} R(k) \cos kp\, dk = \alpha^2(p) + \beta^2(p). \qquad (30.55)$$

If, following the notation of characteristic functions, we write $\phi_R(p)$ for the integral on the left in (30.54) and $\phi_u(p)$ for that on the right in (30.53), we have

$$\phi_R(p) = |\phi_u(p)|^2. \qquad (30.56)$$

We may then put

$$\phi_u(p) = \sqrt{\phi_R}\, e^{i\mu}, \qquad (30.57)$$

where μ is an arbitrary real function. We shall then have

$$\begin{aligned} u(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_u(p)\, e^{-itp}\, dp \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \sqrt{\phi_R} \exp(i\mu - itp)\, dp. \end{aligned} \qquad (30.58)$$

Since $u(t)$ must be real, the imaginary part vanishes and this is equivalent to

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \sqrt{\phi_R} \cos(\mu - tp)\, dp, \qquad (30.59)$$

and μ must be an odd function of p. The result is due to Wiener (1930). It shows that the autocorrelation function R does not uniquely determine $u(t)$ because of the arbitrary function μ.

30.38. Consider now the autocorrelation function $r(k)$ as defined in (30.51). Let us regard the series as defined but equal to zero outside the range $-h$ to $+h$. Then we have

$$2h\, r(k) = \int_{-h}^{h} u(t)\, u(t+k)\, dt = \int_{-\infty}^{\infty} u(t)\, u(t+k)\, dt = R(k), \qquad (30.60)$$

where R and r are zero outside the range $-2h$ to $+2h$. The foregoing results then continue to hold with some modifications concerning factors in 2. If we write—

$$\bar{\phi}_r(p) = \frac{1}{h}\int_{-2h}^{2h} r(k)\, e^{ikp}\, dk = \frac{1}{2h^2}\int_{-\infty}^{\infty} R(k)\, e^{ikp}\, dk \qquad (30.61)$$

and

$$\bar{\phi}_u(p) = \frac{1}{h}\int_{-h}^{h} u(t)\, e^{itp}\, dt = \frac{1}{h}\int_{-\infty}^{\infty} u(t)\, e^{itp}\, dt, \qquad (30.62)$$

then corresponding to (30.56) we have

$$2\,\bar{\phi}_r(p) = |\bar{\phi}_u(p)|^2. \qquad (30.63)$$

We may now let h tend to infinity and observe that the results continue to hold under certain general conditions, provided that the limits exist.

Example 30.7

Consider the series

$$u(t) = A_1 \sin(\lambda_1 t + \alpha_1) + A_2 \sin(\lambda_2 t + \alpha_2) + \ldots + A_m \sin(\lambda_m t + \alpha_m).$$

For the variance we have

$$\lim \frac{1}{2h}\int_{-h}^{h} u^2(t)\, dt = \lim \frac{1}{2h}\int_{-h}^{h} \sum_{j=1}^{m} \{A_j^2 \sin^2(\lambda_j t + \alpha_j)\}\, dt,$$

since the cross-product terms will contribute only a finite amount to the integral and hence vanish in the limit,

$$= \lim \frac{1}{2h} \int_{-h}^{h} \tfrac{1}{2} \Sigma [A_j^2 \{1 - \cos 2 (\lambda_j t + \alpha_j) \}] \, dt$$
$$= \tfrac{1}{2} \Sigma (A_j^2).$$

Similarly for $u(t)\, u(t + k)$ we have

$$\lim \frac{1}{2h} \int_{-h}^{h} [\Sigma \{A_j \sin (\lambda_j t + \alpha_j) \}] [\Sigma \{A_j \sin (\lambda_j t + \lambda_j k + \alpha_j) \}] \, dt$$
$$= \lim \frac{1}{2h} \int_{-h}^{h} \tfrac{1}{2} \Sigma \{A_j^2 [\cos \lambda_j k - \cos \{\lambda_j (2t + k) + 2\alpha_j\}] \} \, dt$$
$$= \tfrac{1}{2} \Sigma A_j^2 \cos \lambda_j k.$$

Thus $\quad r(k) = \dfrac{\Sigma \{A_j^2 \cos (\lambda_j k) \}}{\Sigma A_j^2}.$

The correlogram is the sum of a series of harmonics, like the original series, but the coefficients are different and the harmonics are all in phase.

30.39. The idea underlying the autoregressive scheme of representing time-series may perhaps be best illustrated by an analogy. Imagine a motor-car proceeding along a horizontal road with an irregular surface. The car is fitted with springs which permit it to oscillate to some extent but are designed to damp out the oscillations as soon as the comfort of the passengers will permit. If the car strikes a bump or a pothole in the road the body will oscillate up and down for a time but will soon come to rest so far as vertical motion is concerned. If, however, it proceeds over a continual succession of bumps there will be continual oscillation of varying amplitude and distance between peaks. The oscillations are continually renewed by disturbances, though the distribution of the latter along the road may be quite random. The *regularity* of the motion is determined by the internal structure of the car; but the *existence* of the motion is determined by external impulses.

30.40. It appears to me very plausible to suppose that oscillations in time-series are generated in this way. One does not have to postulate some external rhythmic influence which keeps the oscillation going, or to suppose that the system will oscillate without damping once it has been set in motion. Nor is it necessary to assume that the majority of the deviations between theory and observation are due to "errors" which exert no effect on the subsequent movement of the system. The reader, however, will have to form his own opinion on this matter.* We now proceed to examine an alternative scheme of representation in which the series is represented as a sum of (undamped) cyclic terms.

Periodogram Analysis

30.41. It is well known that under certain general conditions a function $f(t)$ can be expanded in the Fourier series, valid in a certain range,

$$f(t) = a_0 + a_1 \cos \frac{\pi t}{\lambda_1} + a_2 \cos \frac{2\pi t}{\lambda_1} + a_3 \cos \frac{3\pi t}{\lambda_1} + \dots$$
$$+ b_0 + b_1 \sin \frac{\pi t}{\lambda_1} + b_2 \sin \frac{2\pi t}{\lambda_1} + b_3 \sin \frac{3\pi t}{\lambda_1} + \dots \quad . \quad . \quad (30.64)$$

* The scheme considered in this chapter may over-simplify natural conditions in that it assumes finite random disturbances at equidistant time-intervals. If the intervals are not equal, or if the disturbances are small and continually occurring, the autoregressive scheme is only an approximation. Much remains to be done on this subject.

Functions which are not periodic can be expanded in this way; for instance, in the range $0 < x < \pi$,

$$\frac{x}{2} = \sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x - \frac{1}{4}\sin 4x + \ldots$$

The function of course, repeats itself in the range $\pi < x < 2\pi$, and so on.

As a representation of observed series the Fourier series is rather restricted in scope, since the period of every term is a multiple of the fundamental period $2\lambda_1$. A more general scheme is provided by the series

$$f(t) = a_0 + a_1 \cos\frac{2\pi t}{\lambda_1} + a_2 \cos\frac{2\pi t}{\lambda_2} + \ldots$$
$$+ b_0 + b_1 \sin\frac{2\pi t}{\lambda_1} + b_2 \sin\frac{2\pi t}{\lambda_2} + \ldots \quad . \quad . \quad . \quad (30.65)$$

or the alternative form

$$f(t) = A_0 + A_1 \cos\left(\frac{2\pi t}{\lambda_1} + \alpha_1\right) + A_2 \cos\left(\frac{2\pi t}{\lambda_2} + \alpha_2\right) + \ldots \quad . \quad (30.66)$$

Here the λ's are not necessarily commensurable. The object of our analysis is first of all to find out what are the best values of the λ's to select, and secondly to evaluate the other constants a and b, or A and α.

30.42. Suppose we wish to test whether a time-series contains a harmonic term with period μ. Consider the series

$$A = \frac{2}{n}\sum_{j=1}^{n} u_j \cos\frac{2\pi j}{\mu} \quad . \quad . \quad . \quad . \quad (30.67)^*$$

$$B = \frac{2}{n}\sum_{j=1}^{n} u_j \sin\frac{2\pi j}{\mu} \quad . \quad . \quad . \quad . \quad . \quad (30.68)$$

and write

$$S^2 = A^2 + B^2$$
$$= \left|\frac{2}{n}\Sigma\left\{u_j \exp\left(\frac{2\pi j i}{\mu}\right)\right\}\right|^2. \quad . \quad . \quad . \quad . \quad (30.69)$$

Suppose that the series is in fact given by

$$u_j = a \sin\frac{2\pi j}{\lambda} + b_j, \quad . \quad . \quad . \quad . \quad . \quad (30.70)$$

where b_j is a component which we will assume to contain no cyclical element, so that its correlation with the other component is zero, at least for long series. Then we have

$$A = \frac{2a}{n}\sum_{j=1}^{n}\left(\sin\frac{2\pi j}{\lambda}\cos\frac{2\pi j}{\mu}\right) + \frac{2}{n}\sum_{j=1}^{n}\left(b_j \cos\frac{2\pi j}{\mu}\right)$$

* Some writers define these sums with j from 0 to $n - 1$. The signs of A and B may then differ from those given by (30.67) and (30.68), but the intensity and phase are unaffected.

and the second term may be neglected. Thus, writing

$$\alpha = \frac{2\pi}{\lambda}, \qquad \beta = \frac{2\pi}{\mu},$$

we have

$$A = \frac{2a}{n} \Sigma (\sin \alpha j \cos \beta j)$$

$$= \frac{a}{n} \Sigma \{\sin (\alpha - \beta) j + \sin (\alpha + \beta) j\}$$

$$= \frac{a}{n} \left\{ \frac{\sin \frac{1}{2} (\alpha - \beta) n \sin \frac{1}{2} (\alpha - \beta)(n+1)}{\sin \frac{1}{2} (\alpha - \beta)} + \frac{\sin \frac{1}{2} (\alpha + \beta) n \sin \frac{1}{2} (\alpha + \beta)(n+1)}{\sin \frac{1}{2} (\alpha + \beta)} \right\}. \quad . \quad (30.71)$$

For large n this remains small unless α approaches β (or $-\beta$, which is essentially the same situation), and in that case we have

$$A \sim a \sin \tfrac{1}{2} (\alpha - \beta)(n + 1).$$

Similarly,

$$B \sim a \cos \tfrac{1}{2} (\alpha - \beta)(n + 1),$$

so that

$$S^2 = A^2 + B^2 = a^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.72)$$

Thus S remains small unless the "trial" period μ approaches the real period λ, and in that case equals the amplitude a.

30.43. Similarly we may expect that if the series consists of a sum of harmonics with periods $\lambda_1, \lambda_2, \ldots \lambda_m$, S will be small, unless μ is equal to one of these periods, in which case it is finite and equal to the amplitude of the term concerned.

This result forms the basis of what is known as periodogram analysis. We select a number of trial periods for different values of μ and calculate S^2 for each of them. S^2, which is called the *intensity*, is then exhibited as a function of μ, and graphed as ordinate against μ as abscissa. The diagram obtained by joining the points, each to the next, is called the *periodogram*. If this figure has peaks at certain values $\lambda_1 \ldots \lambda_m$ and we are prepared to assume that these are not sampling accidents, the values are the appropriate periods of harmonic terms and the intensity S^2 provides the corresponding amplitudes. The quantities A and B of (30.67) and (30.68) are obtained incidentally and provide the phase angles α of (30.66). We shall illustrate the arithmetic processes below.

30.44. Fig. 30.9 shows the periodogram of the wheat-price index data of Table 30.1. In order not to confuse the diagram for lower values of the trial period we have shown only the major fluctuations. The length of the series was about 300 years from 1545 to 1844, earlier and later figures shown in Table 30.1 not having been taken into account. The primary data have been taken from Sir William Beveridge's classical paper (1922) and are shown in Table 30.9. For practical reasons which will emerge presently, certain trial periods are taken not over exactly 300 years but over the number N of years shown in the table. To reduce the figures to comparability, Beveridge therefore multiplied the sum $A^2 + B^2$ by $\frac{N}{300}$.

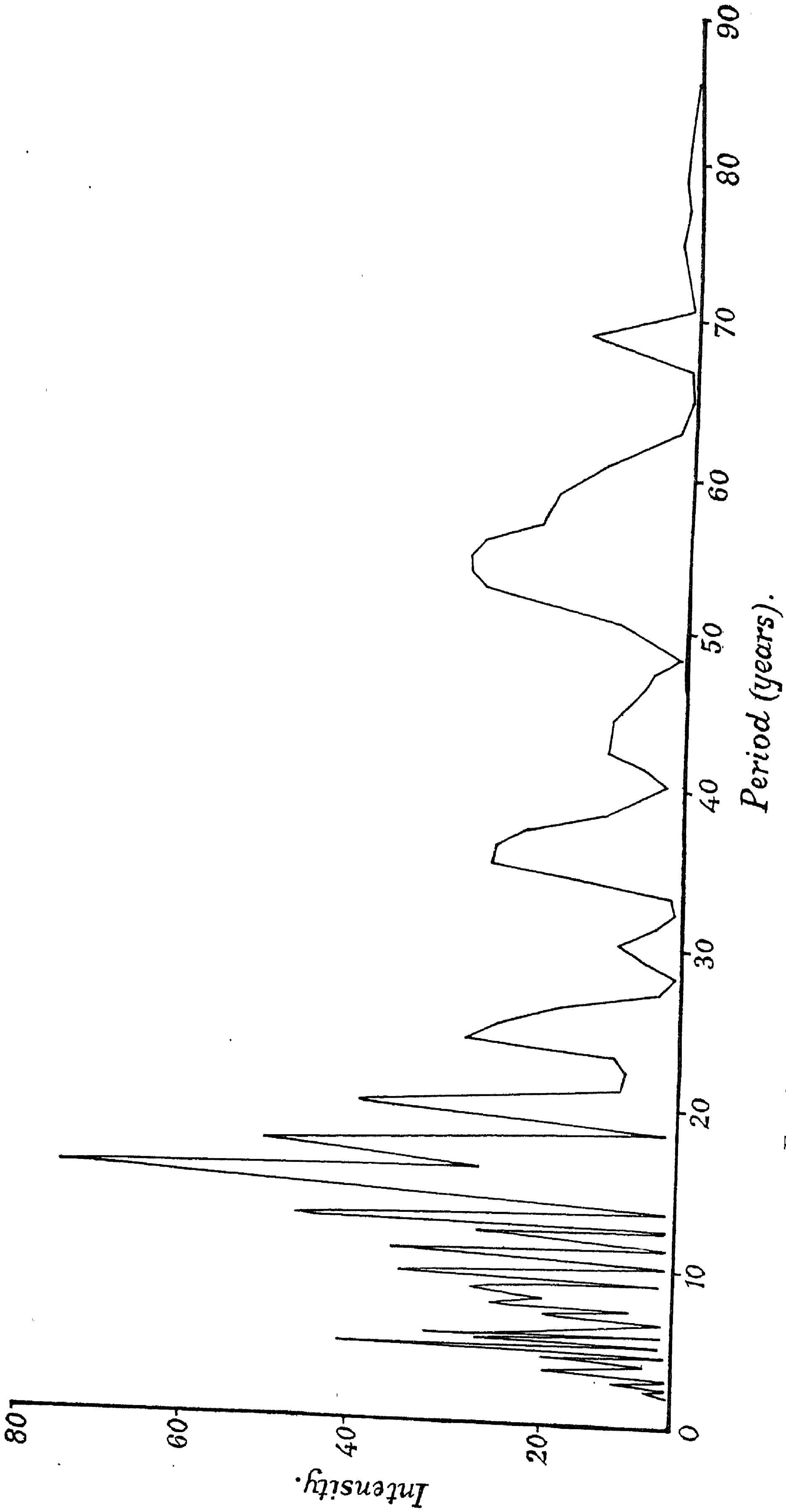


FIG. 30.9.—Periodogram of the Beveridge Wheat-Price Index (Table 30.9).

TABLE 30.9

Periodogram Analysis of the Beveridge Wheat-Price Index Data of Table 30.1.

(From *J.R.S.S.*, 1922, **85**, 412.)

The first observation relates to 1545, except where A and B are given in heavy type.

Period (Years).	Number of Years N.	A.	B.	Intensity $= \frac{N(A^2+B^2)}{300}$.	Period (Years).	Number of Years N.	A.	B.	Intensity $= \frac{N(A^2+B^2)}{300}$.
2·000	300	+ 0·11	—	0·01	2·667	312	− 0·92	+ 1·20	2·38
2·049	336	− **0·40**	− **0·09**	0·19	2·687	301	+ **1·23**	− **0·02**	1·52
2·054	304	+ 0·48	− 0·72	0·77	2·692	315	− 0·04	+ 0·23	0·06
2·061	340	+ **0·38**	− **0·57**	0·54	2·706	322	− 0·27	+ 1·33	1·97
2·069	300	+ 0·25	+ 0·63	0·46	2·714	304	+ 0·83	+ 1·17	2·10
2·074	336	− **0·61**	+ **0·51**	0·71	2·727	300	+ 0·86	+ 1·46	2·87
2·080	312	+ 0·92	− 0·50	1·14	2·733	287	+ 2·05	+ 1·19	6·16
2·087	288	− 0·52	− 0·11	0·27	2·735	279	+ 2·44	+ 1·23	7·82
2·095	308	− 0·91	+ 0·90	1·69	2·737	312	+ 2·23	+ 1·00	6·22
2·105	320	+ 0·90	+ 0·07	0·86	2·741	296	+ 2·43	+ 0·25	5·86
2·112	288	+ 0·90	+ 0·80	1·38	2·750	308	+ 0·90	− 0·84	1·55
2·133	320	+ 0·89	+ 0·15	0·84	2·762	348	− **0·57**	− **0·04**	0·37
2·154	308	+ 0·48	+ 0·23	0·29	2·769	324	+ 1·49	+ 0·23	2·28
2·182	288	+ 1·32	− 0·59	1·99	2·778	325	+ 1·20	− 0·92	2·48
2·200	308	− 0·13	− 0·60	0·39	2·800	336	− **1·01**	− **0·19**	1·18
2·222	320	− 0·32	− 0·62	0·52	2·818	310	+ 0·55	+ 1·07	1·49
2·261	312	+ 0·50	− 0·22	0·31	2·833	323	+ 0·78	− 0·10	0·67
2·286	320	− 0·38	− 0·85	0·93	2·846	296	+ 0·41	+ 0·42	0·34
2·316	308	+ 1·39	− 1·05	3·11	2·857	320	+ 0·96	+ 0·21	1·03
2·333	308	− 0·10	− 0·25	0·08	2·875	322	+ 0·35	+ 0·14	0·15
2.353	320	+ 0·90	+ 0·07	0·86	2·888	312	+ 1·51	+ 0·26	2·43
2.364	312	− 0·12	− 0·63	0·43	2·895	330	− **0·69**	− **1·57**	3·21
2·370	320	+ 0·05	− 0·28	0·08	2·909	320	+ 0·70	− 1·11	1·84
2·375	304	+ 0·29	− 0·43	0·27	2·933	308	− 0·04	+ 0·39	0·16
2·381	300	− 0·19	− 1·22	1·53	2·947	336	− **0·93**	− **1·19**	2·57
2·385	310	− 1·00	− 0·89	1·86	2·960	296	− 0·00	− 1·15	1·30
2·391	330	1·30	0·54	2·18	3·000	300	− 0·29	− 0·39	0·23
2·395	309	− 0·72	+ 0·60	0·90	3·040	304	+ 0·09	+ 0·75	0·58
2·400	312	+ 0·34	+ 0·68	0·60	3·077	320	+ 0·05	+ 1·18	1·50
2·412	328	− 0·08	− 0·65	0·47	3·111	336	+ **0·91**	− **0·44**	1·15
2·417	348	+ 0·63	+ 0·57	0·69	3·143	308	+ 2·01	+ 0·23	4·20
2·435	336	+ 0·44	+ 0·01	0·22	3·167	304	+ 0·46	− 1·05	1·33
2·452	304	− 1·40	− 0·51	2·23	3·200	320	+ 0·43	+ 0·95	1·16
2·462	320	− 0·25	+ 1·49	2·44	3·217	296	+ 1·25	+ 0·00	1·55
2·476	312	− 0·38	+ 0·35	0·27	3·250	312	− 1·22	− 0·47	1·80
2·483	288	− 0·07	+ 0·74	0·53	3·273	324	− 0·55	+ 1·18	1·82
2·500	320	− 0·24	+ 1·19	1·56	3·286	322	− 0·11	+ 0·99	1·07
2·512	324	+ 0·86	+ 0·39	0·97	3·304	304	+ 0·13	+ 0·75	0·59
2·516	312	+ 0·45	+ 0·24	0·26	3·333	320	+ 0·90	+ 1·58	3·54
2·529	301	0·19	− 0·31	0·13	3·364	296	+ 1·76	+ 0·98	4·00
2·545	336	− **1·39**	− **0·81**	2·89	3·375	324	+ 0·55	+ 0·92	1·24
2·555	322	+ 0·38	+ 0·50	0·42	3·385	308	+ 0·35	+ 1·03	1·21
2·571	306	+ 1·25	+ 0·55	1·91	3·400	323	+ 1·12	+ 2·37	7·41
2·588	308	+ 0·30	+ 0·43	0·28	3·407	276	+ 2·98	+ 2·81	14·90
2·600	312	+ 1·02	− 0·39	1·25	3·412	348	+ **1·27**	− **3·98**	15·53
2·615	306	− 0·75	− 0·24	0·63	3·417	328	+ 3·08	− 2·24	15·84
2·625	294	− 0·45	+ 1·36	2·01	3·429	288	+ 3·11	− 1·40	11·16
2·643	296	+ 0·95	− 0·62	1·27	3·444	310	+ 0·09	− 0·99	1·03

TABLE 30.9—*continued.*

Period (Years).	Number of Years N.	A.	B.	Intensity $= \dfrac{N(A^2+B^2)}{300}$.	Period (Years).	Number of Years N.	A.	B.	Intensity $= \dfrac{N(A^2+B^2)}{300}$.
3·455	304	+ 0·55	+ 0·29	0·39	4·933	296	+ 1·57	+ 1·58	4·91
3·462	315	+ 1·57	+ 1·02	4·87	5·000	300	+ 1·85	+ 1·00	4·30
3·500	308	+ 1·20	− 0·94	2·38	5·067	304	− 0·05	+ 3·98	16·09
3·524	296	+ 1·41	− 1·18	3·31	5·091	336	**− 0·73**	**+ 5·55**	35·05
3·538	322	+ 0·50	− 1·45	2·53	5·100	306	+ 5·71	+ 2·98	42·34
3·556	320	+ 0·02	− 0·43	0·20	5·111	322	+ 5·70	+ 0·29	34·91
3·571	325	+ 0·80	− 0·69	1·21	5·125	328	**+ 3·97**	**+ 2·90**	26·38
3·600	324	− 1·03	+ 0·82	1·88	5·143	324	+ 2·46	+ 2·46	13·09
3·619	304	+ 1·18	+ 1·23	2·94	5·200	312	+ 0·02	+ 0·30	0·10
3·636	320	+ 1·14	+ 0·13	1·39	5·250	294	+ 1·74	+ 1·92	6·56
3·643	306	− 0·16	+ 0·27	0·10	5·333	320	+ 0·71	− 4·46	21·72
3·667	308	− 2·14	− 1·07	5·87	5·400	324	+ 1·04	+ 3·71	16·06
3·679	309	+ 0·34	− 1·90	3·83	5·415	325	+ 4·27	+ 1·90	23·66
3·692	288	+ 1·28	− 0·22	1·63	5·429	304	+ 4·72	− 0·28	22·61
3·700	296	+ 0·90	− 0·59	1·18	5·455	300	+ 1·37	− 3·73	15·76
3·714	312	+ 1·15	+ 1·78	4·65	5·500	308	− 1·04	+ 1·49	3·39
3·727	287	− 0·45	− 1·65	2·72	5·555	300	+ 2·40	− 0·68	6·23
3·750	315	+ 0·64	− 0·06	0·44	5·600	336	**+ 0·46**	**+ 1·21**	1·88
3·778	306	− 1·17	− 0·68	1·86	5·667	306	+ 5·31	− 1·97	32·72
3·800	304	+ 1·60	+ 0·80	3·24	5·692	296	+ 2·05	− 3·91	19·18
3·833	322	− 1·12	− 1·63	4·17	5·714	320	+ 0·35	− 2·13	4·97
3·857	324	+ 1·63	+ 0·45	3·08	5·750	322	+ 1·39	− 0·33	2·18
3·888	280	− 0·15	+ 0·66	0·43	5·800	290	+ 3·55	− 2·75	19·47
3·895	296	− 0·66	+ 1·00	1·42	5·846	304	+ 0·00	− 2·29	5·35
3·923	306	+ 0·64	− 1·61	3·06	5·933	356	**+ 4·37**	**+ 0·91**	23·63
3·962	309	− 0·67	+ 1·74	3·59	6·000	300	− 3·50	− 0·12	12·29
4·000	300	+ 1·47	− 1·13	3·64	6·111	330	**− 0·79**	**− 1·90**	4·66
4·077	318	+ 0·57	− 0·26	0·41	6·143	301	+ 0·74	− 2·96	9·32
4·111	296	+ 1·13	− 1·70	4·13	6·167	296	− 0·22	− 2·94	8·56
4·143	290	− 0·50	+ 0·23	0·30	6·200	310	− 2·02	− 3·38	16·02
4·167	325	+ 1·21	+ 0·32	1·70	6·250	325	− 3·23	− 0·11	11·30
4·173	322	+ 0·66	− 1·46	2·77	6·286	308	− 1·72	− 0·59	3·41
4·200	294	− 0·99	− 0·41	1·02	6·333	304	− 1·52	+ 1·29	4·02
4·250	323	+ 0·50	− 2·73	8·32	6·400	320	+ 0·80	+ 2·74	8·71
4·286	300	− 0·65	+ 0·79	1·04	6·500	312	+ 0·69	− 0·73	0·94
4·333	312	− 1·50	− 1·30	4·10	6·571	322	+ 1·49	− 0·77	3·02
4·353	296	− 2·85	− 0·24	8·05	6·667	320	+ 0·25	+ 0·21	0·11
4·364	288	− 2·98	+ 0·75	9·07	6·727	296	+ 0·08	− 0·13	0·02
4·375	315	− 2·47	+ 0·87	7·19	6·750	324	− 0·20	− 1·66	3·01
4·385	342	**− 0·50**	**+ 2·55**	7·72	6·800	306	+ 0·23	− 0·65	0·48
4·400	308	− 1·38	+ 3·27	12·89	6·909	304	+ 0·58	+ 2·56	7·00
4·412	300	+ 0·08	+ 3·62	13·11	6·933	312	+ 1·68	+ 2·01	7·15
4·417	318	+ 0·87	+ 3·85	16·48	7·000	308	+ 3·10	− 2·17	14·74
4·429	310	+ 1·80	+ 2·41	9·32	7·143	300	+ 1·83	− 1·86	6·79
4·444	320	+ 2·15	+ 0·83	5·66	7·200	324	+ 0·54	− 3·93	16·96
4·471	304	+ 0·91	+ 0·79	1·48	7·333	308	+ 1·52	− 2·81	10·46
4·500	306	+ 1·87	+ 0·72	4·09	7·400	296	− 2·33	− 2·72	12·65
4·571	320	− 0·21	+ 0·04	0·22	7·417	356	**+ 1·50**	**− 4·01**	21·72
4·600	322	− 0·08	+ 1·24	1·65	7·429	312	− 3·80	− 1·49	17·28
4·667	336	**+ 0·19**	**+ 0·93**	1·00	7·500	315	+ 0·17	+ 1·50	2·40
4·750	304	− 0·12	+ 2·28	5·28	7·600	304	− 2·33	− 1·37	7·43
4·800	288	+ 2·44	+ 1·08	6·84	7·667	322	− 1·46	− 2·61	9·57
4·857	306	− 1·06	− 1·30	2·89	7·750	310	+ 1·38	− 0·39	2·13
4·888	312	− 1·80	+ 2·11	8·00	7·857	330	**− 0·50**	**+ 0·28**	0·36

TABLE 30.9—*continued.*

Period (Years).	Number of Years N.	A.	B.	Intensity $= \frac{N(A^2+B^2)}{300}$.	Period (Years).	Number of Years N.	A.	B.	Intensity $= \frac{N(A^2+B^2)}{300}$.
8·000	312	− 3·96	+ 1·34	18·67	17·500	280	− 6·18	− 4·45	54·12
8·091	356	**+ 4·32**	**− 0·98**	23·23	18·000	306	− 4·40	+ 1·25	21·29
8·200	287	+ 1·62	− 0·64	2·90	18·500	296	− 1·46	+ 2·25	7·10
8·222	296	+ 0·19	− 0·56	0·34	19·000	304	+ 1·00	− 0·23	1·07
8·333	325	+ 0·21	+ 0·91	0·95	19·750	316	− 4·73	− 1·59	26·25
8·500	323	+ 0·17	+ 3·19	10·41	20·000	320	− 5·71	+ 1·69	37·88
8·667	312	+ 2·51	− 1·01	7·59	21·000	294	+ 0·78	+ 2·61	7·28
8·800	308	+ 2·97	+ 0·83	9·77	22·000	308	+ 1·87	+ 1·58	6·18
9·000	306	− 1·51	− 0·57	2·65	23·000	322	− 2·45	− 1·43	8·61
9·200	322	− 0·16	− 1·56	2·65	24·000	288	+ 0·45	+ 5·19	26·10
9·333	336	**− 0·74**	**+ 0·64**	1·08	24·667	296	+ 4·31	+ 1·99	22·21
9·500	304	+ 1·08	+ 1·07	2·26	25·000	325	+ 3·86	− 0·19	14·94
9·667	290	+ 5·03	+ 0·37	24·55	26·000	312	+ 1·23	− 1·34	3·43
9·750	312	+ 4·46	− 3·56	33·89	27·000	324	+ 0·50	− 0·33	0·38
9·818	324	+ 1·21	− 4·94	27·90	28·000	308	− 0·49	+ 0·68	0·72
10·000	320	− 1·19	− 0·83	2·25	29·000	290	+ 1·08	− 2·12	5·46
10·200	306	+ 0·86	− 0·22	0·80	30·000	300	− 1·53	− 2·34	7·81
10·250	328	**− 0·69**	**+ 1·10**	1·84	31·000	310	− 1·98	+ 0·13	4·06
10·400	312	+ 1·88	− 1·65	6·52	32·000	320	− 0·37	+ 0·51	0·42
10·500	294	+ 2·46	− 1·82	9·19	33·000	330	**+ 0·96**	**− 0·78**	1·68
10·750	301	+ 1·47	− 3·13	11·98	34·000	306	− 3·00	− 2·15	13·90
10·800	324	+ 1·00	− 4·75	25·48	35·000	280	− 4·64	+ 1·79	23·11
11·000	308	− 3·85	− 4·26	33·84	36·000	288	− 1·65	+ 4·85	23·29
11·200	336	**− 2·48**	**+ 0·55**	7·24	37·000	296	+ 2·08	+ 3·92	19·47
11·500	322	− 1·32	− 0·66	2·34	38·000	304	+ 2·99	+ 0·56	9·37
11·667	280	+ 0·46	+ 1·42	2·07	40·000	320	− 1·44	− 0·63	2·63
12·000	312	− 2·47	− 4·04	23·30	41·000	328	**− 1·93**	**+ 0·93**	5·01
12·143	340	**− 0·22**	**− 4·37**	21·66	42·000	294	+ 0·93	+ 3·02	9·75
12·333	296	− 2·44	+ 2·74	11·43	44·000	308	+ 3·00	− 0·14	9·27
12·500	325	− 1·22	+ 2·63	9·13	45·000	315	+ 1·69	− 1·99	7·14
12·667	304	+ 2·28	+ 5·19	32·58	46·000	322	+ 0·16	− 2·27	5·58
12·800	320	+ 5·70	+ 3·26	46·01	48·000	288	− 0·76	− 0·09	0·56
12·875	309	+ 6·46	+ 0·77	43·58	50·000	300	+ 1·83	+ 2·19	8·14
13·000	312	+ 4·26	− 4·32	38·23	52·000	312	+ 4·77	− 0·57	24·03
13·333	320	+ 0·40	+ 0·37	0·32	53·000	318	+ 4·22	− 2·60	26·08
13·500	324	+ 2·56	− 2·09	11·79	54·000	324	+ 2·84	− 4·01	26·09
13·667	328	**+ 3·49**	**− 1·34**	15·28	55·000	330	+ 3·54	− 3·30	25·82
14·000	308	+ 1·15	− 1·00	2·38	56·000	336	**+ 3·31**	**− 2·36**	18·47
14·500	290	− 3·78	− 0·18	13·82	58·000	290	+ 3·89	+ 1·49	16·82
14·667	308	− 1·50	+ 4·23	20·69	60·000	300	− 3·08	− 0·93	10·32
15·000	300	+ 6·32	− 2·66	46·83	62·000	310	− 1·62	+ 0·39	2·88
15·200	304	+ 1·19	− 8·52	75·04	64·000	320	− 0·78	+ 0·13	0·66
15·250	305	− 0·28	− 8·65	76·17	66·000	330	− 0·56	− 0·56	0·69
15·286	321	− 2·35	− 7·15	60·62	68·000	340	**+ 2·90**	**− 1·88**	13·58
15·333	322	− 3·89	− 6·55	62·29	70·000	280	− 0·69	− 0·16	0·47
15·500	310	− 6·92	− 2·02	59·11	74·000	296	− 1·20	+ 0·82	2·07
16·000	320	− 1·46	+ 4·52	24·02	76·000	304	− 0·66	+ 1·17	1·83
16·667	300	+ 5·21	− 0·39	27·33	78·000	312	+ 0·58	+ 1·26	2·00
17·000	306	+ 2·56	− 6·35	47·84	80·000	320	+ 0·77	+ 0·82	1·34
17·333	312	− 3·04	− 6·65	54·55	84·000	336	**+ 0·26**	**+ 0·69**	0·62

An examination of the periodogram suggests the possibility of 20 periods, as follows :—

Period (Years).	Corrected Intensity $\frac{N(A^2+B^2)}{300}$.	Period (Years).	Corrected Intensity $\frac{N(A^2+B^2)}{300}$.
2·735	7·82	11·000	33·84
3·417	15·84	12·000	23·30
4·417	16·48	12·800	46·01
5·100	42·34	15·250	76·17
5·415	23·66	17·333	54·55
5·667	32·72	20·000	37·88
5·933	23·63	24·000	26·10
7·417	21·72	35·000	23·29
8·091	23·23	54·000	26·09
9·750	33·89	68·000	13·58

This is evidently rather an embarrassing profusion of possibilities, and we cannot immediately accept all these periods as significant. Sir William discussed them in detail in the original paper and was inclined to attribute reality to 18 or 19 of them, partly on grounds which do not concern us here, such as the existence of weather oscillations with these "periods". In particular, where a period had a high intensity he analysed the two halves of the series separately to see whether the periods persisted, finding that most of them did.

30.45. An inspection of the correlogram of the series in Fig. 30.5 reveals a striking difference between the two methods of analysis. From the correlogram we should be inclined to suspect a mean period of about 15 years, corresponding to the peak of greatest intensity in the periodogram, with a subsidiary ripple of about 5 to 6 years' period, corresponding to one or more of the peaks in the periodogram ; but of the other 18 periods there is no sign. The conclusion is inevitable that either the correlogram is insensitive or the periodogram is misleading. Having raised this highly important question we shall, unfortunately, have to leave it unsettled in part ; but we shall show that at least three-quarters of the periods thrown up for consideration by the periodogram are not significant.

30.46. The calculation of the intensity S^2 depends on that of the quantities A and B of equations (30.67) and (30.68). Suppose in the first place that our trial period μ is an integer. We then write down the series in rows of μ, thus :—

$$\left.\begin{array}{lllll} u_1 & u_2 & u_3 & \cdots & u_\mu \\ u_{\mu+1} & u_{\mu+2} & u_{\mu+3} & \cdots & u_{2\mu} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ u_{(\rho-1)\mu+1} & u_{(\rho-1)\mu+2} & u_{(\rho-1)\mu+3} & \cdots & u_{\rho\mu} \\ \hline \text{Totals } m_1 & m_2 & m_3 & \cdots & m_\mu \end{array}\right\} \quad . \quad (30.73)$$

We continue writing down the rows until there are fewer than μ terms remaining, the extra terms being left out of account. The number $\rho\mu$ is then as near in multiples of μ as we can get to the number in the series n, and may be denoted by N. This array is sometimes known as the Buys-Ballot table.

We then form the sum—

$$\frac{2}{\rho\mu}\left\{m_1 \cos\frac{2\pi}{\mu} + m_2 \cos\frac{4\pi}{\mu} + \ldots + m_\mu \cos\frac{2\mu\pi}{\mu}\right\} \quad . \quad . \quad (30.74)$$

and this is clearly the quantity A of (30.67) for the series of N terms. Similarly we have

$$B = \frac{2}{\rho\mu}\sum_{j=1}^{\mu}\left(m_j \sin\frac{2\pi j}{\mu}\right). \quad . \quad . \quad . \quad . \quad (30.75)$$

If the trial period μ is a rational fraction $\frac{\nu}{\sigma}$ we write the series down in rows of ν and proceed in the same way; and if it is irrational or is a number which gives a large value of ν when expressed as a fraction, we take two convenient neighbouring values of μ and interpolate in the periodogram.

30.47. In actual practice we do not write down the array (30.73). The sums m may be formed on an adding machine by starting with u_1 and then adding every μth member to give m_1; then starting with u_2 and adding every μth member to give m_2, and so on. Or alternatively, the values may be written on cards, one for each member of the series, and the pack dealt into μ heaps. The total of the m's, together with any members left over, equals the sum of the series and provides a check on the work.

Example 30.8

Consider the Beveridge series of Table 30.1. For the trial period 2 we may take 300 terms of the series, and m_1' (about zero mean) will be the sum of the values $u_1, u_3 \ldots u_{299}$ and m_2' will be the sum of the values with even subscripts. These sums are for the years 1545 to 1844 inclusive,

$$m_1' = 14{,}909$$
$$m_2' = 14{,}893.$$

The mean is 14,901, so that about the mean of the series

$$m_1 = +8$$
$$m_2 = -8.$$

Now, for a trial period 2, $\sin\frac{2\pi j}{2}$ vanishes and hence $B = 0$. For A we have (in our notation, which gives different signs from Beveridge's to A and B)—

$$A = \frac{2}{300}\left\{m_1 \cos\frac{2\pi}{2} + m_2 \cos\frac{4\pi}{2}\right\}$$
$$= \frac{2}{300}\{m_2 - m_1\}$$
$$= -\frac{32}{300} = -0{\cdot}11.$$

Thus $$S^2\,(\text{corrected}) = \frac{300}{300}A^2 = 0{\cdot}01,$$

as shown in Table 30.9.

For a trial period 2·600, we could take $\mu = \frac{13}{5}$ and arrange the series in rows of 13, requiring 23 rows accounting for 299 values of the series. We may, however, save ourselves some arithmetic by taking 24 rows, a multiple of 4, occupying 312 observations.

Or rather, we take 6 rows of 52, giving us the values for a trial period 52; then add m_1 to m_{27}, m_2 to m_{28} and so on, giving the result we would have got by taking 12 rows of 26 and hence providing the values for a trial period of 26; then we add again in the same way, and so on, obtaining successively the values of m required for trial periods of 13, 6·5, and 3·25. Similarly, by multiplying the original 52 values of m by the respective values of $\cos \frac{20\pi j}{52}$ and $\sin \frac{20\pi j}{52}$ we get the values of A and B required for a trial period of $\frac{52}{10}$. It is thus evident that we can use the single set of 52 values of m to provide the required constants for trial periods $\frac{52}{1}, \frac{52}{2}, \frac{52}{3}$, and so forth. This is the main reason why, in Table 30.9, 312 observations are shown as N for the trial periods 2·080, 2·261, 2·364, 2·476, 2·600, 2·737, 2·888, 3·250, 3·714, 4·333, 5·200, 6·500, 7·429, 8·667, 10·400, 13·000, 17·333, 26·000 and 52·000. The arithmetic, though difficult enough, is not as laborious as appears at first sight.

30.48. There is an interesting relation between the periodogram and the correlogram by which the latter, in theory, determines the former. We consider, as in **30.38**, a function $u(t)$ defined at every point of time in some range $-h$ to h. Then

$$\begin{aligned}\bar{\alpha}(p) + i\bar{\beta}(p) &= \frac{1}{h}\int_{-h}^{h} e^{ipt}\, u(t)\, dt \\ &= \frac{1}{h}\int_{-h}^{h} \cos pt\, u(t)\, dt + \frac{i}{h}\int_{-h}^{h} \sin pt\, u(t)\, dt \quad . \quad . \quad (30.76)\end{aligned}$$

corresponds to the sums of (30.67) and (30.68) and may be written $A + iB$, where

$$p = \frac{2\pi}{\mu}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.77)$$

It follows that the intensity S^2 is related to the Fourier transform of $r(k)$ by the relation, derived from (30.63),

$$\begin{aligned}S^2 &= 2\bar{\phi}_r(p) \\ &= \frac{2}{h}\int_{-2h}^{2h} r(k)\, e^{ikp}\, dk, \quad . \quad . \quad . \quad (30.78)\end{aligned}$$

which is true also in the limit, subject to conditions of existence. Thus the intensity is, if $r(k)$ exists over an infinite range, the quantity—

$$\lim \frac{2}{h}\int_{-2h}^{2h} r(k)\cos kp\, dk,$$

and if $R(k)$ exists the parallel quantity—

$$\int_{-\infty}^{\infty} R(k)\cos kp\, dk.$$

The periodogram is thus derivable from the autocorrelation function. Since the latter does not uniquely determine the series the periodogram will not do so either.

Example 30.9

Consider the autocorrelation function, which in present notation may be written

$$R(k) = \frac{\rho^k \sin(k\theta + \psi)}{\sin \psi}.$$

This, as we have seen, represents the correlogram of an autoregressive series of the simple linear kind involving u_{t+2}, u_{t+1} and u_t. We may write this as

$$R(k) = \frac{e^{-qk}\sin(k\theta + \psi)}{\sin\psi}, \qquad q > 0$$

since p is less than unity. It is to be remembered that since $R(-k) = R(k)$, the modulus of k is to be used when k is negative.

We have

$$S^2 = \int_{-\infty}^{\infty} \frac{e^{-|qk|}\sin(k\theta + \psi)}{\sin\psi}\cos kp\, dk$$

$$= \int_{-\infty}^{\infty} e^{-|qk|}\cos k\theta \cos kp\, dk$$

$$= \frac{q}{q^2 + (\theta + p)^2} + \frac{q}{q^2 + (\theta - p)^2}.$$

This is the intensity in the periodogram of the series, p being the quantity $\frac{2\pi}{\mu}$ and not to be confused with our original damping factor p.

It is remarkable that, as μ becomes large, S^2 tends to the constant value $\frac{2q}{q^2 + \theta^2}$, that is to say, the periodogram tends to a fixed level, without peaks. From the analogy with the analysis of light-rays into colours (each colour corresponding to a particular harmonic), we may say that the periodogram develops a "continuous spectrum". In a very interesting chapter on periodogram analysis Davis (1941) has given a number of examples exhibiting this kind of effect.

Significance of a Periodogram

30.49. Suppose that the values $u_1 \ldots u_n$ are random elements from a normal population with variance σ^2. Then the function

$$A = \frac{2}{n}\sum_{j=1}^{n} u_j \cos\frac{2\pi j}{\mu}$$

is normally distributed with variance

$$\text{var } A = \frac{4\sigma^2}{n^2}\sum_{j=1}^{n}\cos^2\frac{2\pi j}{\mu}$$

$$= \frac{2\sigma^2}{n}; \qquad \ldots \ldots \quad (30.79)$$

and similarly

$$\text{var } B = \frac{2\sigma^2}{n}. \qquad \ldots \ldots \quad (30.80)$$

We also see that cov $(A, B) = 0$ so that A and B are independent. Hence the joint distribution of A and B is

$$dF = \frac{n}{4\pi\sigma^2}\exp\left\{-\frac{n}{4\sigma^2}(A^2 + B^2)\right\} dA\, dB. \qquad \ldots \quad (30.81)$$

Thus the distribution of $S^2 = A^2 + B^2$ is

$$dF = \frac{n}{4\sigma^2} \exp\left(-\frac{n}{4\sigma^2} S^2\right) dS^2. \qquad . \quad . \quad . \quad . \quad . \quad (30.82)$$

The probability that S^2 exceeds $\frac{4\sigma^2\kappa}{n}$ in value is immediately obtainable as $e^{-\kappa}$.

30.50. This result is due to Schuster (1898), but it gives only the probability that a value of S^2 *chosen at random* will exceed a given value; whereas in the periodogram we deliberately pick out the biggest values for inspection. Walker (1914) pointed out that if $e^{-\kappa}$ is small the probability that all of m independent values of S^2 should not exceed $\frac{4\sigma^2\kappa}{n}$ is $(1 - e^{-\kappa})^m$, so the probability that at least one should exceed that amount is

$$1 - (1 - e^{-\kappa})^m. \qquad . \quad . \quad . \quad . \quad . \quad (30.83)$$

Davis (1941) gives tables of this function.

30.51. Both the Schuster and the Walker tests depend on a knowledge of σ^2. Since the mean value of S^2 in (30.82) is $\frac{4\sigma^2}{n}$, the usual procedure is to consider the test as a comparison of S^2 with $E(S^2)$; but σ^2 itself has to be estimated from the original data.

30.52. Fisher (1929*a*) has given a test which avoids the inexactitude due to the estimation of σ^2. If v is the estimate and S^2 is the *largest* intensity, then the probability that

$$g = \frac{S^2}{2v} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (30.84)$$

will exceed a given value is

$$\nu(1-g)^{\nu-1} - \binom{\nu}{2}(1-2g)^{\nu-1} + \ldots + (-1)^{m-1}\binom{\nu}{m}(1-mg)^{\nu-1}, \quad (30.85)$$

where $\nu = \frac{1}{2}(n-1)$, n being the (odd) number of observations, and m is the greatest integer less than $1/g$. The result was extended by Stevens (1939*a*)—see also Fisher (1940*a*) and Finney (1941*a*). Davis (1941) also gives tables of this function.

30.53. All the tests we have described are based on random normal variation in the original series; but in practice nobody would embark on the labour of a periodogram analysis unless he had satisfied himself that the data were not random. It seems to me, therefore, that these tests are really off the main point, being tests based on a hypothesis which we have already rejected. They are not without their usefulness, however. We may assume with some confidence that if a particular intensity in the series is not shown as significant on the hypothesis of random variation, it is not significant when the series is systematic. What does not follow is that if one intensity is significant then others must be so, even if they exceed the significance values; for they are not independent of the significant value, at least for short series. What we ought to do, perhaps, is to extract the component which is considered significant from the series and then analyse the remainder; and so on as long as significant terms appear. But this is hardly a practical computational possibility. Tests of significance in the periodogram, as in the correlogram, remain undiscovered.

Example 30.10

Let us examine the significance of the 20 periods of the Beveridge periodogram given in **30.44.**

Sir William gave the value of $\frac{4\sigma^2}{n}$ in his original paper as 5·898. Expressing the intensities as a multiple κ of this amount, we find :—

Period.	κ.	Period.	κ.
2·735	1·33	11·000	5·74
3·417	2·69	12·000	3·95
4·417	2·79	12·800	7·80
5·100	7·18	15·250	12·91
5·415	4·01	17·333	9·25
5·667	5·55	20·000	6·42
5·933	4·01	24·000	4·43
7·417	3·68	35·000	3·95
8·091	3·94	54·000	4·42
9·750	5·75	68·000	2·30

There are 305 trial periods in Table 30.9. Let us consider the probability that at least one of 305 *independent* values of κ will exceed given values, that is to say, the probabilities given by (30.83). We find—

κ	Probability.
2	1·000
4	0·996
6	0·531
8	0·097
10	0·014

On this basis we should be inclined to attribute significance to the period 15·25, for which $\kappa = 12{\cdot}91$. We have no right to be surprised that at least one value exceeds $\kappa = 6$. If we take this value as the critical one, only the periods 5·100, 12·800, 15·250, 17·333 and 20·000 would be significant, that is to say, five out of 20.

Again, since $e^{-5} = 0{\cdot}007$, we should expect to find in 305 independent members two in excess of 5. Actually there are eight. But they are not independent and we cannot rely on this comparison to say that six are significant. On the whole, however, it looks as if at least three-quarters of the periods are not significant, and possibly more. The example will illustrate the difficulty of testing the significance of the periodogram as a whole.

Lag Correlation

30.54. The idea of serial correlation can be extended to the joint variation of two series. If we have two series $u(t)$, $v(t)$ in standard measure, we may define the lag correlation of order k as

$$r(k) = \int u(t)\, v(t+k)\, dt, \qquad . \quad . \quad . \quad . \quad (30.86)$$

where the integral includes summation in the case when the series are specified at equidistant points of time. We note that in this case $r(k)$ is not equal to $r(-k)$ and $r(0)$ is not unity.

Table 30.10 shows the lag correlations between two series of English wheat prices and horse populations (for the original series see Kendall, 1944*a*). The data are shown as a lag correlogram in Fig. 30.10.

TABLE 30.10

Lag Correlations for Two Series of English Wheat Prices and Horse Populations (Deviations from a Simple Nine-Year Average).

(The order of the correlation is the number of years by which horse population lags behind wheat price, e.g. r_{10} is the correlation of wheat price with the horse population of ten years earlier.)

Order of Correlation k.	r_k.	Order of Correlation k.	r_k.
− 10	− 0·22	1	− 0·24
− 9	− 0·19	2	− 0·36
− 8	− 0·24	3	− 0·12
− 7	− 0·16	4	0·16
− 6	− 0·09	5	0·17
− 5	0·07	6	0·39
− 4	0·27	7	0·36
− 3	0·31	8	0·15
− 2	0·41	9	− 0·16
− 1	0·25	10	− 0·44
0	− 0·12		

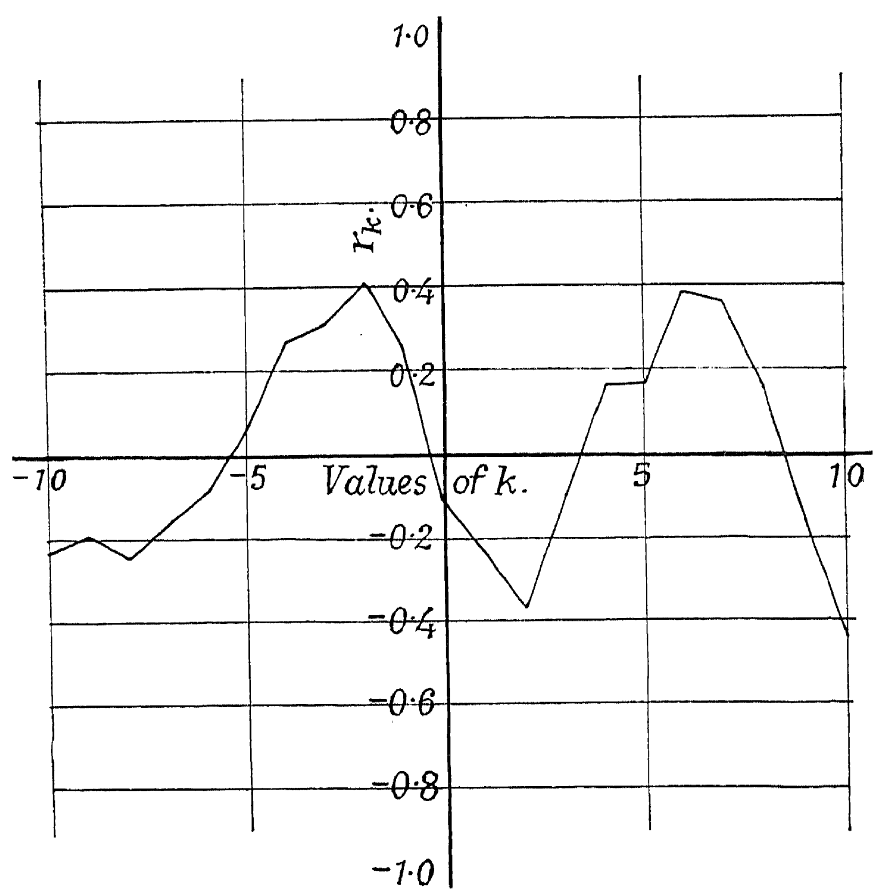


FIG. 30.10.—Lag Correlation of Wheat Prices and Horse Populations (Table 30.10).

The systematic appearance is unmistakable and we notice in particular that the maximum correlation occurs between the wheat price and the horse population of two years later. This bears the obvious explanation that when a farmer earns more he buys or breeds more horses ; but it does not follow logically that this must be so or that there need be any causal nexus between the two series. If two autoregressive series are oscillating with mean periods which are close together and only a short span of experience is available for scrutiny, then lag correlations of the damped sinusoidal type may appear, as it were, by accident.

30.55. We have now reached the end of our account of the statistical analysis of time-series and the end of this book ; and the final words we have to say of the one will apply generally to the other. Much has been left unsaid, partly from lack of space, partly from deficiencies in the present state of knowledge, and partly from a desire not to overburden the reader. We have not avoided mathematical analysis where it was necessary to advance the argument ; but we have insisted on the expression of results in numerical form and the necessity of experimental confirmation whenever it could be obtained. That there are gaps in the treatment we have given and unexplored branches of the subject to which we have barely referred are not entirely matters of regret ; for the over-early and peremptory reduction of knowledge into arts and methods is one of the errors which Bacon cautioned us against more than 300 years ago. Much remains to be done ; and this book will have served its purpose if the reader is left with the desire to do some of it himself.

NOTES AND REFERENCES

The theoretical aspects of the autoregressive series and of moving averages are discussed in Wold's book on *The Analysis of Stationary Time-Series* (1938*a*). The basic memoir is that by Yule (1927*a*) on sunspots. For applications to meteorology see Walker (1931) and to economics Kendall (1944*a*). Davis's book on *The Analysis of Economic Time Series* (1941) contains a great deal of interesting material but should not be read uncritically. Two earlier papers by Yule (1921 and 1926) are also of interest. See also my paper on " The Analysis of Oscillatory Time-Series " in the *Journal of the Royal Statistical Society* for 1945, a paper by Yule in the same journal, my brochure (in press) on " Researches in Oscillatory Time-Series ", and a symposium introduced by Bartlett in the *Supplement* to the Journal for 1946.

The classical work on periodogram analysis is that of Schuster (1898). The books by Brunt (1931) on *The Combination of Observations* and by Whittaker and Robinson (1940) on *The Calculus of Observations* contain useful introductory accounts ; and Davis's book referred to above has an excellent chapter illustrated with an unusual number of examples. Papers by Crum (1923) and Greenstein (1935) are of interest. The papers by Sir William Beveridge (1921, 1922) on wheat prices and rainfall have been justly described by Davis as a heroic piece of periodogram analysis. Tables facilitating the calculation of intensities were published by Turner (1913), and more complete tables will be given in my brochure referred to above. See also the book by Stumpff (1937).

Various short-cut methods of periodogram analysis have been proposed by several authors, e.g. Oppenheim (1909), Bruns (1921) and Alter (1933, 1937) ; but their value is problematical. There is a useful memoir by Bartels (1935) which is worth studying.

EXERCISES

30.1. For the autoregressive series

$$u_{t+2} + au_{t+1} + bu_t = \varepsilon_{t+2}$$

show that if ε is a random variable and the series is long,

$$\frac{\text{var } u}{\text{var } \varepsilon} = \frac{1 + b}{(1 - b)\,\{(1 + b)^2 - a^2\}},$$

and hence that the variance of the generated series may be much greater than that of ε itself.

30.2. For the autoregressive series of the previous exercise use the relation

$$r_{k+2} + ar_{k+1} + br_k = 0, \qquad k \geqslant -1$$

to derive the relation

$$r_k = \frac{p^k \sin (k\theta + \psi)}{\sin \psi}.$$

30.3. If the estimated coefficients a' and b' in the autoregressive scheme are reduced in the manner of **30.32** by a superposed error, show that

$$\frac{b'}{a'} > \frac{b}{a}.$$

(Yule, 1927*a*.)

30.4. Show that if, in the autoregressive scheme of Exercise 30.1, $b = 1$, the series becomes undamped and the correlogram reduces to a simple harmonic. Examine the effect on the solution (30.23).

30.5. If any series has fitted to it a series generated by the scheme of Exercise 30.1, a and b being any constants, show that for the serial correlations of the residuals, say σ_k, we have

$$\sigma_k = \frac{(1 + a^2 + b^2)\,\rho_k + a\,(1 + b)\,(\rho_{k+1} + \rho_{k-1}) + b\,(\rho_{k+2} + \rho_{k-2})}{1 + a^2 + b^2 + 2a\,(1 + b)\,\rho_1 + 2b\rho_2}.$$

30.6. Show that the series with an autocorrelation function

$$r\,(k) = \frac{\sin \lambda k}{\lambda k}$$

has a periodogram which is zero for periods less than $\frac{\pi}{\lambda}$ and has ordinate $\frac{\pi}{\lambda}$ for periods greater than $\frac{\pi}{\lambda}$, i.e. has a continuous spectrum.

30.7. In equation (30.71), noting that the dominant term vanishes for $\alpha - \beta = \frac{2m\pi}{n}$, where m is an integer, show that for such a "vanishing" trial period μ

$$\mu = \lambda\left(1 + \frac{m}{n}\mu\right), \text{ approximately.}$$

Hence the width of a peak in the periodogram is approximately $\frac{2\lambda^2}{n}$, and the main peak will be flanked by smaller peaks of the same width. (This " side-band " effect is another complication in the interpretation of the periodogram, but not apparently a very serious one.)

30.8. If a series of values $u_1 \dots u_n$ is supplemented by a number of zeros as $u_0, u_{-1}, u_{-2} \dots u_{n+1}, u_{n+2}$, etc., as far as is necessary, and the resulting series differenced, show that

$$\tau_j = P_0\binom{2j}{j} - 2P_1\binom{2j}{j-1} + 2P_2\binom{2j}{j-2} - \dots + 2(-1)^j P_j,$$

where τ_j is the sum of squares of jth differences and $P_j = \sum_{k=1}^{n-j} x_k x_{k+j}$. Hence show that the arithmetic of serial correlation may be related to that of the variate-difference method, and vice-versa.

30.9. Show that the serial correlations of a long series obtained by differencing a random series m times are given by

$$r(k) = (-1)^k \frac{m(m-1)\dots(m-k+1)}{(m+1)\dots(m+k)}$$

and hence that the correlogram of such a series oscillates.

(Yule, 1921.)

30.10. The Whittaker periodogram. Writing

$$\eta^2(\mu) = \frac{\text{var } m}{\text{var } u},$$

where var u is the variance of the series and var m is the variance of the sums m of (30.73), show that if

$$u_j = a \sin \frac{2\pi j}{\lambda} + b_j,$$

where b_j is uncorrelated with periodic terms, then

$$\eta^2(\mu) = \frac{\dfrac{a^2 \mu^2 \sin^2 \frac{N\pi}{\lambda}}{2N^2 \sin^2 \frac{\mu\pi}{\lambda}} + \dfrac{\mu}{N} \text{var } b}{\frac{1}{2}a^2 + \text{var } b}.$$

Hence show that, in the neighbourhood of λ, the graph of η as ordinate with μ as abscissa (Whittaker's periodogram) has a peak of breadth $\frac{2\lambda^2}{N}$ flanked by smaller peaks.

(Whittaker, *Month. Notes R. Astr. Soc.*, 1911, **71**; cf. Whittaker and Robinson, *Calculus of Observations*.)

APPENDIX A

ADDENDA TO VOLUME I

(1) *Frequency and Distribution Functions*

An interesting paper by Burr (1942) considers the possibility of fitting elementary mathematical functions, not to the *frequency* function as has been the almost universal practice hitherto, but direct to the *distribution* function. This approach seems to merit further attention. In general, the distribution function has fewer analytical peculiarities than the frequency function—for instance, it cannot be infinite—and in applications to sampling it is the former which is nearly always required. The frequency function can, of course, be derived from the distribution function to a close approximation by differencing, or differentiation, processes which are usually easier to carry out than the inverse processes of integration.

(2) *Extension of the Carleman Criterion* (**4.22**)

Cramér and Wold (1936) have extended Carleman's criterion for uniqueness in the problem of moments in the following form :—

If

$$\lambda_i = \mu'_{i00\ldots} + \mu'_{0i0\ldots} + \mu'_{00i\ldots} + \ldots .$$

the distribution is completely determined by its moments if

$$\Sigma \left\{ \frac{1}{(\lambda_{2m})^{\frac{1}{2m}}} \right\}$$

diverges. It is rather interesting that the criterion is independent of the product-moments.

(3) *Convergence of Series Leading to Standard Errors*

The usual type of expansion in differentials, exemplified in **9.6**, raises a point of mathematical difficulty in that the differentials themselves and the remainder terms, though usually small, may sometimes be large for sampling reasons, however large the sample. The necessary rigorisation of the process has been given by Derkson (1939) in terms of the notion of *stochastic convergence*, that is to say, a sort of statistical convergence in which the series converges nearly always in a precisely defined sense.

(4) *Moments of Moments for Finite Populations*

The formulae for moments of the mean and variance in samples from a finite population were stated without proof in **11.26**. It is obvious that if in these results we let N, the population number, tend to infinity, we obtain the formulae for sampling from an infinite population. Irwin and I (1944) have recently shown that the process may be reversed and the formulae for the finite case derived from those for the infinite case. This offers the simplest and most direct method of deriving the formulae known to me. Reference may also be made to Sukhatme, " On Bipartitional Functions " (*Phil. Trans.*, 1938, A, **237**, 375) and " Moments and Product-Moments of Moment-statistics for Samples of the Finite and Infinite Populations " (*Sankhyā*, 1944, **6**, 363).

(5) *Tied Ranks*

In the treatment of rank correlation in Chapter 16 it was assumed that ranking was always possible; but in practice cases occur when two or more individuals "tie" and the ranks have to be equalised in some way. This possibility introduces the most intractable complications into theoretical work, but sometimes ties occur so frequently that a systematic method of dealing with them is necessary. The subject has been reviewed and reconsidered by Woodbury (1940) and more recently by myself (*Biom.*, 1945, **33**, part 3).

(6) *Coefficients of Rank Correlation*

Daniels (1944) has recently unified the theory of rank correlation by showing that Spearman's ρ, my τ and the product-moment coefficient are particular cases of a general coefficient. In particular he has demonstrated the formula for the covariance of ρ and τ given in **16.24** as very probably true.

APPENDIX B

BIBLIOGRAPHY

The following Bibliography has no pretensions to completeness in spite of its length. It contains about half the titles recorded in my own notes, which themselves are doubtless far from comprehensive. Nevertheless, I hope it will be useful to those readers who want to take their studies of particular subjects somewhat further. By consulting the references given here and following up the references which they themselves provide, it should be possible for the reader to acquaint himself with most of what is known, or at least with what is worth knowing, about a particular topic.

The names of authors are not included in the Index (pages 504 ff.) unless they occur in the text, since the Bibliography itself is arranged alphabetically under authors' names. The subjects, however, are indexed, and anyone wishing to consult references on a particular topic should refer in the first place to the Index, which in turn will refer to the authors who have dealt with the matter in question.

In general the Bibliography contains only references to theoretical papers; applications and illustrative material are included only when some theoretical point is involved. Papers which have been superseded by later work are omitted, except where they have a historical interest.

In compiling this material I have been particularly indebted to the valuable periodical reviews of Recent Advances in Mathematical Statistics by Irwin, Hartley and others in the *Journal of the Royal Statistical Society:* 1932, **95**, 498; 1934, **97**, 114; 1935, **98**, 88; 1936, **99**, 714; 1938, **101**, 394; 1939, **102**, 406; and 1940, **103**, 534.

Many papers written since 1939 are included, but some journals are not available in war-time so that foreign work published after the entry of various countries into the war may be incompletely represented. Where possible, the references have been checked against the original publications, but here also I have had to rely on second-hand references in cases where the original papers were inaccessible.

Note.—Names beginning with de, del, le, St., van, von, etc., are entered under those titles, i.e. the order is strictly alphabetical.

ABERNETHY, J. R. (1933). On the elimination of systematic errors due to grouping. *Ann. Math. Stats.*, **4**, 263.

ACKERMANN, W. G. (1939). Eine Erweiterung des Poissonschen Grenzwertsatzes und ihre Anwendung auf die Risikoprobleme in der Sachversicherung. *Schrift. math. Inst. Berlin*, **4**, 211.

ADCOCK, R. J. (1878). A problem in least squares. *Analyst*, **5**, 53.

AITKEN, A. C., and OPPENHEIM, A. (1931). On Charlier's new form of the frequency function. *Proc. Roy. Soc. Edin.*, **51**, 35.

AITKEN, A. C. (1931). Some applications of generating functions to normal frequency. *Quart. J. Maths.*, **2**, 130.

AITKEN, A. C. (1932). On the orthogonal polynomials in frequencies of Type B. *Proc. Roy. Soc. Edin.*, **52**, 174.

AITKEN, A. C. (1933*a*). On the graduation of data by the orthogonal polynomials of least squares. *Proc. Roy. Soc. Edin.*, **53**, 54.

AITKEN, A. C. (1933*b*, *c*). On fitting polynomials to weighted data by least squares. *Proc. Roy. Soc. Edin.*, **54**, 1; and: On fitting polynomials to data with weighted and correlated errors. *Ibid.*, **54**, 12.

AITKEN, A. C. (1935*a*). On least squares and linear combination of observations. *Proc. Roy. Soc. Edin.*, **55**, 42.

AITKEN, A. C., and GONIN, H. T. (1935*b*). On fourfold sampling with and without replacement. *Proc. Roy. Soc. Edin.*, **55**, 114.

AITKEN, A. C. (1937*a*, *b*, 1938). Studies in practical mathematics: I. The evaluation with applications of a certain triple product matrix. *Proc. Roy. Soc. Edin.*, **57**, 172; II. The evaluation of the latent roots and latent vectors of a matrix. *Ibid.*, **57**, 269; III. The application of quadratic extrapolation to the evaluation of derivatives and to inverse interpolation. *Ibid.*, **58**, 161.

AITKEN, A. C., and SILVERSTONE, H. (1942). On the estimation of statistical parameters. *Proc. Roy. Soc. Edin.*, **61**, 186.

ALLAN, F. E. (1930). The general form of the orthogonal polynomials for simple series with proofs of their simple properties. *Proc. Roy. Soc. Edin.*, **50**, 310.

ALLAN, F. E., and WISHART, J. (1930). A method of estimating the yield of a missing plot in field experimentation work. *J. Agr. Sci.*, **20**, 399.

ALLEN, H. V. (1938). A theorem concerning the linearity of regression. *Stat. Res. Mem.*, **2**, 60.

ALLEN, R. G. D. (1939). The assumptions of linear regression. *Economica*, **6**, 191.

ALT, F. L. (1942). Distributed lags. *Econometrika*, **10**, 113.

ALTER, D. (1924). Application of Schuster's periodogram to long rainfall records, beginning 1748. *Monthly Weather Rev.*, **52**, 479.

ALTER, D. (1925). Equations extending Schuster's periodogram. *Astr. J.*, **36**, No. 850.

ALTER, D. (1926*a*). An examination by means of Schuster's periodogram of rainfall data from long records in typical sections of the world. *Monthly Weather Rev.*, **54**, 44.

ALTER, D. (1926*b*). The criteria of reality in the periodogram. *Monthly Weather Rev.*, **54**, 57.

ALTER, D. (1933). An extremely simple form of periodogram analysis. *Proc. Nat. Acad. Sci.*, **19**, 335.

ALTER, D. (1937). A simple form of periodogram. *Ann. Math. Stats.*, **8**, 121.

ALTER, D. (1939). Correction of sample moment bias due to lack of high contact and to histogram grouping. *Ann. Math. Stats.*, **10**, 192.

'ALUMNUS' (1932). A comparison of the effect of rainfall on spring and autumn-dressed wheat at Rothamsted Experimental Station, Harpenden. *J. Agr. Sci.*, **22**, 101.

AMBARZUMIAN, G. (1937). Verteilungskurven der Wahrscheinlichkeiten, welche im Limit die Verteilungskurven von Pearson ergeben. *C.R. Acad. Sci. U.S.S.R.*, **16**, 251.

ANDERSON, O. (1914). Nochmals über 'The elimination of spurious correlation due to position in time and space.' *Biom.*, **10**, 269.

ANDERSON, O. (1923). Über ein neues Verfahren bei Anwendung der 'Variate-difference' Methode. *Biom.*, **15**, 134. Corrigenda, **15**, 423.

ANDERSON, O. (1926). Über die Anwendung der Differenzenmethode (Variate-difference method) bei Reihenausgleichungen, Stabilitätsuntersuchungen, und Korrelationsmessungen. *Biom.*, **18**, 293.

ANDERSON, O. (1927). On the logic of the decomposition of statistical series into separate components. *J.R.S.S.*, **90**, 548.

ANDERSON, O. (1929). *Die Korrelationsrechnung in der Konjunkturforschung.* Schroeder, Bonn.

ANDERSON, O. (1935). *Einführung in die mathematische Statistik.* Springer, Wien.

ANDERSON, P. H. (1942). Distributions in stratified sampling. *Ann. Math. Stats.*, **13**, 42.

ANDERSON, R. L. (1942). Distribution of the serial correlation coefficient. *Ann. Math. Stats.*, **13**, 1.

ANDERSON, T. F. (1935). Some further notes upon experiments with actuarial functions and Fourier's series. *J. Inst. Act.*, **67**, 31.

ANDERSSON, W. (1932). Researches into the theory of regression. *Medd. Lunds Astr. Obs.*, Series 2, No. 64.

ANDERSSON, W. (1934). On a new method of computing non-linear regression curves. *Ann. Math. Stats.*, **5**, 81.

ANDRÉ, D. (1884). Étude sur les maxima, minima et séquences des permutations. *Ann. Éc. Norm. Sup.*, (3), **1**, 121.

AROIAN, L. A. (1937). The Type B Gram-Charlier Series. *Ann. Math. Stats.*, **8**, 183.

AROIAN, L. A. (1941). A study of R. A. Fisher's z-distribution and the related F-distribution. *Ann. Math. Stats.*, **12**, 429.

AROIAN, L. A. (1943). A new approximation to the levels of significance of the chi-square distribution. *Ann. Math. Stats.*, **14**, 93.

AUMANN, G. (1934–1935). Aufbau von Mittelwerten mehrere Argumente. *Math. Ann.*, **109**, 235, and **111**, 713.

AYYANGAR, A. A. K. (1934). Note on the recurrence formulae for the moments of the point binomial. *Biom.*, **26**, 262; and: Note on the incomplete moments of the hypergeometrical series. *Ibid.*, **26**, 264.

AYYANGAR, A. A. K. (1938). On the semi-invariants of two variates and their additive property. *Sankhyā*, **4**, 85, and *J. Indian Math. Soc.*, **3**, 1.

BACON, H. M. (1938). Note on a formula for the multiple correlation coefficient. *Ann. Math. Stats.*, **9**, 227.

BAILEY, A. L. (1931). The analysis of covariance. *J. Am. Stat. Ass.*, **26**, 424.

BAKER, G. A. (1930*a*). Transformations of bimodal distributions. *Ann. Math. Stats.*, **1**, 334.

BAKER, G. A. (1930*b*). The significance of the product-moment coefficient, with special reference to the marginal distributions. *J. Am. Stat. Ass.*, **25**, 387.

BAKER, G. A. (1930*c*). Random samples from non-homogeneous populations. *Metron*, **8**, No. 3, 67.

BAKER, G. A. (1930*d*). Distribution of the means of samples of n drawn at random from a population represented by the Gram-Charlier Series. *Ann. Math. Stats.*, **1**, 199.

BAKER, G. A. (1931). The relation between the means and variances, means squared and variances in samples from combinations of normal populations. *Ann. Math. Stats.*, **2**, 333.

BAKER, G. A. (1932). Distribution of the means divided by the standard deviations of samples from non-homogeneous populations. *Ann. Math. Stats.*, **3**, 1.

BAKER, G. A. (1934). Transformation of non-normal frequency-distributions into normal distributions. *Ann. Math. Stats.*, **5**, 113.

BAKER, G. A. (1935). Note on the distribution of the standard deviation and second moments from a Gram-Charlier distribution. *Ann. Math. Stats.*, **6**, 127.

BAKER, G. A. (1936). The probability that the mean of a second sample will differ from the mean of a first sample by less than a certain multiple of the standard deviation of the first sample. *Ann. Math. Stats.*, **7**, 197.

BAKER, G. A. (1937). Correlation surfaces of two or more indices when the components of the indices are normally distributed. *Ann. Math. Stats.*, **8**, 179.

BAKER, G. A. (1938). The probability that the standard deviation of a second sample will differ from the standard deviation of a first sample by a certain multiple of the first sample. *Metron*, **13**, No. 3, 49.

BAKER, G. A. (1940). A comparison of Pearsonian approximations with exact sampling distributions of means and variances. *Ann. Math. Stats.*, **11**, 219.

BAKER, G. A. (1941). Tests of homogeneity for normal populations. *Ann. Math. Stats.*, **12**, 233.

BARBACKI, S., and FISHER, R. A. (1936). A test of the supposed precision of systematic arrangements. *Ann. Eug. Lond.*, **7**, 189.

BARNARD, M. M. (1935). The secular variations of skull characters in four series of Egyptian skulls. *Ann. Eug. Lond.*, **6**, 352.

BARNARD, M. M. (1936). An enumeration of the confounded arrangements in the $2 \times 2 \times 2$ factorial designs. *Supp. J.R.S.S.*, **3**, 195.

BARTELS, J. (1935). Zur Morphologie geophysikalischer Zeitfunktionen. *Sitz. Berl. Akad. Wiss.*, **139**.

BARTKY, W. (1943). Multiple sampling with constant probability. *Ann. Math. Stats.*, **14**, 363.

BARTLETT, M. S. (1933*a*). On the theory of statistical regression. *Proc. Roy. Soc. Edin.*, **53**, 260.

BARTLETT, M. S. (1933*b*). Probability and chance in the theory of statistics. *Proc. Roy. Soc.*, **A**, **141**, 518.

BARTLETT, M. S. (1934*a*). The problem in statistics of testing several variances. *Proc. Camb. Phil. Soc.*, **30**, 164.

BARTLETT, M. S. (1934*b*). The vector representation of a sample. *Proc. Camb. Phil. Soc.*, **30**, 327.

BARTLETT, M. S. (1935*a*). The effect of non-normality on the *t*-distribution. *Proc. Camb. Phil. Soc.*, **31**, 223.

BARTLETT, M. S. (1935*b*). Contingency table interactions. *Supp. J.R.S.S.*, **2**, 248.

BARTLETT, M. S. (1935*c*). Some aspects of the time-correlation problem in regard to tests of significance. *J.R.S.S.*, **98**, 536.

BARTLETT, M. S. (1935*d*). An examination of the value of covariance in dairy-cow nutrition experiments. *J. Agr. Sci.*, **25**, 238.

BARTLETT, M. S. (1936*a*). The information available in small samples. *Proc. Camb. Phil. Soc.*, **32**, 560.

BARTLETT, M. S. (1936*b*). Statistical information and properties of sufficiency. *Proc. Roy. Soc.*, **A**, **154**, 124.

BARTLETT, M. S. (1936*c*). A note on the analysis of covariance. *J. Agr. Sci.*, **26**, 488.

BARTLETT, M. S. (1936*d*). Square-root transformations in the analysis of variance. *Supp. J.R.S.S.*, **3**, 68.

BARTLETT, M. S. (1936*e*). Some notes on insecticide tests in the laboratory and in the field. *Supp. J.R.S.S.*, **3**, 185.

BARTLETT, M. S. (1937*a*). Sub-sampling for attributes. *Supp. J.R.S.S.*, **4**, 131.

BARTLETT, M. S. (1937*b*). Note on the derivation of fluctuation formulae for statistical assemblies. *Proc. Camb. Phil. Soc.*, **33**, 390.

BARTLETT, M. S. (1937*c*). Properties of sufficiency and statistical tests. *Proc. Roy. Soc.*, **A**, **160**, 268.

BARTLETT, M. S. (1937*d*). Some examples of statistical methods of research in agriculture and applied biology. *Supp. J.R.S.S.*, **4**, 137.

BARTLETT, M. S. (1937*e*). The statistical conception of mental factors. *Brit. J. Psych.*, **28**, 97.

BARTLETT, M. S. (1938*a*). The approximate recovery of information from replicated experiments with large blocks. *J. Agr. Sci.*, **28**, 418.

BARTLETT, M. S. (1938*b*). The characteristic function of a conditional statistic. *J. Lond. Math. Soc.*, **13**, 62.

BARTLETT, M. S. (1938*c*). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.*, **34**, 33.

BARTLETT, M. S. (1939*a*). Complete simultaneous fiducial distributions. *Ann. Math. Stats.*, **10**, 129.

BARTLETT, M. S. (1939*b*). A note on tests of significance in multivariate analysis. *Proc. Camb. Phil. Soc.*, **35**, 180.

BARTLETT, M. S. (1939*c*). The standard errors of discriminant function coefficients. *Supp. J.R.S.S.*, **6**, 169.

BARTLETT, M. S. (1940). A note on the interpretation of quasi-sufficiency. *Biom.*, **31**, 391.

BARTLETT, M. S. (1941). The statistical significance of canonical correlations. *Biom.*, **32**, 29.

BATEN, W. D. (1931). Corrections for the moments of a frequency distribution in two variables. *Ann. Math. Stats.*, **2**, 309.

BATEN, W. D. (1933*a*). Frequency laws for the sum of *n* variables which are subject to given frequency laws. *Metron*, **10**, No. 3, 75.

BATEN, W. D. (1933*b*). Sampling from many parent populations. *Tokohu Math. Journ.*, **36**, 206.

BATEN, W. D. (1934). The probability law for the sum of n independent variables, each subject to the law $(1/2h)$ sech $(\pi x/2h)$. *Bull. Am. Math. Soc.*, **40**, 284.

BATTIN, I. L. (1942). On the problem of multiple matching. *Ann. Math. Stats.*, **13**, 294.

BAYES, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans.*, **53**, 370.

BEALE, F. S. (1937). On the polynomials related to the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} \equiv \frac{N}{D}.$$

Ann. Math. Stats., **8**, 206.

BEALL, G. (1939). Methods of estimating the population of insects in a field. *Biom.*, **30**, 422.

BEALL, G. (1942). The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biom.*, **32**, 243.

BECK, E. (1936). Existenzbeweise zur Wahrscheinlichkeitstheorie. *Math. Zeit.*, **41**, 222.

BECKER, R., PLANT, H., and RUNGE, I. (1930). *Anwendung der mathematischen Statistik auf Probleme der Massenfabrikation.* Springer, Berlin.

BEHRENS, W. V. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.*, **68**, 807.

BELARDINELLI, G. (1934). Su una teoria astratta del calcolo della probabilità. *Giorn. Ist. Ital. Att.*, **5**, 418.

BENINI, R. (1926). *Principi di statistica metodologica.* Unione Tipografica Editrice Torinese, Turin.

BENNETT, T. L. (1920). The theory of measurement of changes in the cost of living. *J.R.S.S.*, **83**, 455.

BERGE, P. O. (1938). A note on a form of Tchebycheff's theorem for two variables. *Biom.*, **29**, 405.

BERGSTRÖM, S. (1918). Sur les moments de la fonction de correlation normale de n variables. *Biom.*, **12**, 177.

BERKSON, J. (1930). Bayes' theorem. *Ann. Math. Stats.*, **1**, 42.

BERKSON, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *J. Am. Stat. Ass.*, **33**, 526.

BERNOUILLI, J. (1713). *Ars coniectandi.* (A German translation in Ostwald's *Klassiker der Exakten Wissenschaften*, Nos. 107 and 108.)

BERNSTEIN, F. (1932). Die mittleren Fehlerquadrate und Korrelation der Potenzmomente und ihre Anwendung auf Funktionen der Potenzmomente. *Metron*, **10**, No. 3, 3.

BERNSTEIN, F. (1937). Regression and correlation evaluated by a method of partial sums. *Ann. Math. Stats.*, **8**, 77.

BERNSTEIN, S. (1927). Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.*, **97**, 1.

BERNSTEIN, S. (1936). Détermination d'une limite inférieure de la dispersion des sommes de grandeurs liées en chaîne singulière. *Rec. Math. Moscou*, **1**, 29.

BERNSTEIN, S. (1937). Sur quelques modifications de l'inégalité de Tchebycheff. *C.R. Acad. Sci. U.S.S.R.*, **17**, 279.

BERTRAND, J. L. F. (1889). *Calcul des probabilités.* Gauthier-Villars, Paris.

BESICOVITCH, A. S. (1932). *Almost Periodic Functions.* Cambridge University Press.

BESSON, L. (trans. and abridged by E. W. Woolard) (1920). On the comparison of meteorological data with chance results. *Monthly Weather Rev.*, **48**, 89.

BEVERIDGE, Sir W. H. (1921). Weather and harvest cycles. *Econ. J.*, **31**, 429.

BEVERIDGE, Sir W. H. (1922). Wheat prices and rainfall in Western Europe. *J.R.S.S.*, **85**, 412.

BHATTACHARYA, D. P., and NARAYAN, R. D. (1942). Moments of D^2-statistic for populations with unequal dispersions. *Sankhyā*, **5**, 401.

BHATTACHARYA, K. N. (1943). A note on twofold triple systems. *Sankhyā*, **6**, 313.

BILHAM, E. G. (1926). Correlation coefficients. *Q. J. Roy. Met. Soc.*, **52**, 172.

BINGHAM, M. D. (1941). A new method for obtaining the inverse matrix. *J. Am. Stat. Ass.*, **36**, 530.

BISHOP, D. J. (1939). On a comprehensive test for the homogeneity of variances and covariances in multivariate problems. *Biom.*, **31**, 31.

BISHOP, D. J., and NAIR, U. S. (1939). A note on certain methods of testing for the homogeneity of a set of estimated variances. *Supp. J.R.S.S.*, **6**, 89.

BISPHAM, J. W. (1922). Note on a heterotypic frequency function. *J.R.S.S.*, **85**, 488.

BISPHAM, J. W. (1920, 1923). An experimental determination of the distribution of the partial correlation coefficient in samples of thirty. *Proc. Roy. Soc.*, **A**, **97**, 1920, and *Metron*, **2**, 684, 1923.

BLAKEMAN, J. (1905). On tests for linearity of regression in frequency-distributions. *Biom.*, **4**, 332.

BLAKEMAN, J., and PEARSON, K. (1906). On the probable error of the coefficient of mean square contingency. *Biom.*, **5**, 191.

BLISS, C. I. (1935). The calculation of the dosage-mortality curve. *Ann. App. Biol.*, **22**, 134; and: The comparison of dosage-mortality data. *Ibid.*, **22**, 307.

BLISS, C. I. (1937). The calculation of the time-mortality curve. *Ann. App. Biol.*, **24**, 816.

BLISS, C. I. (1938). The transformation of percentages for use in the analysis of variance. *Ohio J. Sci.*, **38**, 9.

BLÜMEL, H. (1939). Bemerkungen über die Sheppardsche Korrektur. *Arch. math. Wirtsch.- u. Sozialforschung*, **5**, 39.

BOAS, R. P., and SMITHIES, F. (1937). On the characterisation of a distribution function by its Fourier transform. *Am. J. Maths.*, **60**, 513.

BOCHNER, S., and JESSEN, B. (1934). Distribution functions and positive definite functions. *Ann. Maths.*, **35**, 252.

BOCHNER, S. (1936). A converse of Poisson's theorem in the theory of probability. *Ann. Maths.*, **37**, 816.

BOCHNER, S. (1937). Stable laws of probability and completely monotone functions. *Duke Math. J.*, **3**, 726.

BÖDEWADT, G. T. (1936). Zum Momentproblem für das Intervall [0, 1]. *Math. Zeit.*, **40**, 426.

BOHR, H. (1925). Zur theorie fast periodische Funktionen. *Acta Math.*, **45**, 29.

BONFERRONI, C. (1933). Sulla probabilità massima nello Schema di Poisson. *Giorn. Ist. Ital. Att.*, **4**, 109.

BONFERRONI, C. (1939). Di una estensione del coefficiente di correlazione. *Giorn. degli Economisti*, Nov.-Dec., p. 7.

BOREL, E. (editor) (1925 and subsequently). *Traité du calcul des probabilités et de ses applications.* Gauthier-Villars, Paris.

BOREL, E. (1933). Sur un problème élémentaire de probabilités et la quasi-périodicité de certains phénomènes arithmétiques. *Comptes rendus*, **196**, 881.

BOREL, E. (1937). Sur l'imitation du hasard. *Comptes rendus*, **204**, 203.

BOREL, E. (1939). Sur une interprétation des probabilités virtuelles. *Comptes rendus*, **208**, 1369; and: Sur certains problèmes de répartition et les probabilités virtuelles. *Ibid.*, **208**, 1177.

BOSE, A. N. (1941). Some problems of field operations in labour inquiries. *Sankhyā*, **5**, 229.

BOSE, C. (1943). Note on the sampling error in the method of double sampling. *Sankhyā*, **6**, 329.

BOSE, R. C. (1934). On the application of hyperspace geometry to the theory of multiple correlation. *Sankhyā*, **1**, 338.

BOSE, R. C. (1936*a*). On the exact distribution and moment-coefficients of the D^2-statistic. *Sankhyā*, **2**, 143.

BOSE, R. C. (1936*b*). A note on the distribution of differences in mean values of two samples

drawn from two multivariate normally-distributed populations, and the definition of the D^2-statistic. *Sankhyā*, **2**, 379.

Bose, R. C. (1938*a*). On the distribution of the means of samples drawn from a Bessel function population. *Sankhyā*, **3**, 262.

Bose, R. C. (1938*b*). On the application of Galois fields to the problem of construction of Hyper-Graeco-Latin squares. *Sankhyā*, **3**, 323.

Bose, R. C., and Roy, S. N. (1938*c*). The distribution of the studentised D^2-statistic. *Sankhyā*, **4**, 19.

Bose, R. C. (1939). On the construction of balanced incomplete block designs. *Ann. Eug. Lond.*, **9**, 353.

Bose, R. C., and Nair, K. R. (1939). Partially balanced incomplete block designs. *Sankhyā*, **4**, 337.

Bose, R. C., and Roy, S. N. (1940). The use and distribution of the studentised D^2-statistic when the variances and covariances are based on k samples. *Sankhyā*, **4**, 535.

Bose, R. C., and Kishen, K. K. (1941). On the problem of confounding in general symmetrical factorial designs. *Sankhyā*, **5**, 21.

Bose, R. C. (1942*a*). A note on the resolvability of balanced incomplete block designs. *Sankhyā*, **6**, 105.

Bose, R. C., and Nair, K. R. (1942*b*). On complete sets of Latin squares. *Sankhyā*, **5**, 361.

Bose, S. N. (1935). On the complete moment coefficients of the D^2-statistic. *Sankhyā*, **2**, 385.

Bose, S. N. (1937). On the moment coefficients of the D^2-statistic and certain integral and differential equations connected with the multivariate normal population. *Sankhyā*, **3**, 105.

Bose, S. S. (1934*a*). Tables for testing the significance of linear regression in the case of time-series and other single-valued samples. *Sankhyā*, **1**, 277.

Bose, S. S. (1934*b*). A note on the mathematical expectation of the value of the regression coefficient. *Sankhyā*, **1**, 432.

Bose, S. S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population. *Sankhyā*, **2**, 65.

Bose, S. S. (1938*a*). On a Bessel function population. *Sankhyā*, **3**, 253.

Bose, S. S. (1938*b*). Relative efficiency of regression coefficients estimated by the method of finite differences. *Sankhyā*, **3**, 339.

Bose, S. S., and Mahalanobis, P. C. (1938*a*). On the exact test of association between the occurrence of thunderstorm and abnormal ionisation. *Sankhyā*, **3**, 249.

Bose, S. S., and Mahalanobis, P. C. (1938*b*). On estimating individual yields in the case of mixed-up yields of two or more plots in field experiments. *Sankhyā*, **4**, 103.

Bowley, A. L. (1912). The measurement of the accuracy of an average. *J.R.S.S.*, **75**, 77.

Bowley, A. L. (1919). The measurement of changes in the cost of living. *J.R.S.S.*, **82**, 343.

Bowley, A. L. (1920). *Prices and Wages in the United Kingdom*, Clarendon Press, Oxford.

Bowley, A. L., and Smith, K. C. (1924). *Seasonal variations in Finance, Prices and Industry*. Lond. and Camb. Ec. Service, Special Memo. No. 7.

Bowley, A. L. (1925). Measurement of the precision attained in sampling. *Bull. Int. Inst. Stat.*, **22**, 1er livre.

Bowley, A. L. (1926). The influence on the precision of index-numbers of the correlation between the prices of commodities. *J.R.S.S.*, **89**, 300.

Bowley, A. L. (1928). *F. Y. Edgeworth's Contributions to Mathematical Statistics*. Royal Statistical Society, London.

Bowley, A. L. (1933). The action of economic forces in producing frequency-distributions of income, prices and other phenomena. *Econometrika*, **1**, 358.

Bowley, A. L. (1938). Note on Professor Frisch's 'The Problem of Index Numbers'. *Econometrika*, **6**, 83.

BRADLEY, P. D., and CRUM, W. L. (1939). Periodicity as an explanation of variation in hog production. *Econometrika*, **7**, 221.

BRADY, J. (1935). A biological application of the analysis of covariance. *Supp. J.R.S.S.*, **2**, 99.

BRANDER, F. A. (1933). A test of the significance of the difference of the correlation coefficients in normal samples. *Biom.*, **25**, 102.

BRANDT, A. E. (1933). The analysis of variance in a 2 $\times$ s table with disproportionate frequencies. *J. Am. Stat. Ass.*, **28**, 164.

BRELOT, M. (1936, 1937). Sur l'influence des erreurs de mesure en statistique. *J. Math. Pur. App.*, **15**, 113, and **16**, 285 ; also *Congrès Int. de Math.*, Oslo (1936).

BRELOT, M. (1937). Quelques difficultés dans l'application pratique de la théorie des erreurs. *Mathematica*, **13**, 243.

BRODERICK, P. S. (1937). On some symbolic formulae in probability theory. *Proc. Roy. Irish Acad.*, **A**, **44**, 19.

BROGGI, U. (1934). Su di uno speciale problema dei momenti. *Ann. di Mat.*, (4), **12**, **63**.

BROWN, G. M. (1933). On sampling from compound populations. *Ann. Math. Stats.*, **4**, **288**.

BROWN, G. W. (1939). On the power of the L_1 test for equality of several variances. *Ann. Math. Stats.*, **10**, 119.

BROWN, G. W. (1940). Reduction of a certain class of statistical hypotheses. *Ann. Math. Stats.*, **11**, 254.

BROWN, J. W., GREENWOOD, M., and WOOD, Frances (1914). A study of index correlations. *J.R.S.S.*, **77**, 317.

BROWN, W. (1909). Some experimental results in correlation. *Proceedings Sixth Int. Congress Psychology, Geneva.*

BROWN, W., and THOMSON, G. H. (1925). *The essentials of mental measurement.* Cambridge University Press.

BROWN, W. (1935). A note on the theory of two factors versus the sampling theory of mental ability. *Brit. J. Psych.*, **25**, 395.

BROWNLEE, J. (1905). Statistical studies in immunity : small-pox and vaccination. *Biom.*, **4**, 313.

BROWNLEE, J. (1910). The significance of the correlation coefficient when applied to Mendelian distributions. *Proc. Roy. Soc. Edin.*, **30**, 473.

BROWNLEE, J. (1911). The mathematical theory of random migration and epidemic distribution. *Proc. Roy. Soc. Edin.*, **31**, 262.

BROWNLEE, J., and MORISON, R. M. (1911). Notes on the calculation of the probabilities of life at high ages. *J.R.S.S.*, **74**, 201.

BROWNLEE, J. (1918). Certain aspects of the theory of epidemiology in special reference to plague. *Proc. Roy. Soc. Med., Sect. Epidem. and State Medicine*, **10D**, 85.

BROWNLEE, J. (1924*a*). Experiments to test the theory of goodness of fit. *J.R.S.S.*, **87**, **76**.

BROWNLEE, J. (1924*b*). Test of periodogram analysis. *J.R.S.S.*, **87**, **83**.

BROWNLEE, J. (1925). Error in the correlation due to random sampling when proportionate mortalities are used. *J.R.S.S.*, **88**, 105.

BRUEN, C. (1938). Methods for the combination of observations, etc. *Metron*, **13**, No. 2, 61.

BRUNS, H. (1906). *Wahrscheinlichkeitsrechnung und Kollektivmasslehre.* Teubner, Leipzig.

BRUNS, H. (1921). Über die Analyse periodischer Vorgänge. *Astr. Nach.*, **188**.

BRUNT, D. (1925). Periodicities in European weather. *Phil. Trans.*, **A**, **225**, 247.

BRUNT, D. (1928). Harmonic analysis and the interpretation of the results of periodogram investigations. *Mem. R. Met. Soc.*, **2**, No. 15, 47.

BRUNT, D. (1931). *The Combination of Observations.* Cambridge University Press.

BUNAK, V. V. (1936). Changes in the mean values of characters in mixed populations. *Ann. Eug. Lond.*, **7**, 195.

BURKHARDT, F., and STACKELBERG, H. V. (1939). Zur Ableitung der Sheppardschen Korrektur. *Arch. Math. Wirtsch.- u. Socialforschung*, **5**, 127.

BURKS, B. S. (1933). A statistical method for estimating the distribution of sizes of completed fraternities in a population represented by a random sampling of individuals. *J. Am. Stat. Ass.*, **28**, 388.

BURNSIDE, W. (1924). On Bayes' formula. *Biom.*, **16**, 189.

BURNSIDE, W. (1928). *Theory of Probability.* Cambridge University Press.

BURR, I. W. (1942). Cumulative frequency functions. *Ann. Math. Stats.*, **13**, 215.

BURRAU, C. (1934). Contribution to the problem of dissection of a given frequency curve. *Nordic Stat. J.*, **5**, 43.

BURT, C. (1927). *Mental and Scholastic Tests.* P. S. King, London.

BURT, C. (1936). *Marks of Examiners.* Macmillan, London.

BURT, C. (1937*a*). Correlations between persons. *Brit. J. Psych.*, **28**, 59.

BURT, C. (1937*b*). Methods of factor analysis with and without successive approximations. *Brit. J. Educ. Psych.*, **7**, 172.

BURT, C. (1938*a*). The unit hierarchy and its properties. *Psychometrika*, **3**, 151.

BURT, C. (1938*b*). Factor analysis by sub-matrices. *J. Psych.*, **6**, 339.

BUYS-BALLOT, C. H. D. (1847). *Les changements périodiques de température.* Utrecht.

CACCIOPPOLLI, R. (1932). Sull' approssimazione per polinomi delle funzioni definiti in campi illimitati. *Giorn. Ital. Ist. Att.*, **3**, 364.

CAMP, B. H. (1922). A new generalisation of Tchebycheff's statistical inequality. *Bull. Am. Math. Soc.*, **28**, 427.

CAMP, B. H. (1924). Probability integrals for the point binomial. *Biom.*, **16**, 163.

CAMP, B. H. (1925*a*). Probability integrals for the hypergeometric series. *Biom.*, **17**, 61.

CAMP, B. H. (1925*b*). Mutually consistent multiple regression surfaces. *Biom.*, **17**, 443.

CAMP, B. H. (1932). The converse of Spearman's two-factor theorem. *Biom.*, **24**, 418.

CAMP, B. H. (1934). Spearman's general factor again. *Biom.*, **26**, 260.

CAMP, B. H. (1937). Methods of obtaining probability distributions. *Ann. Math. Stats.*, **8**, 90.

CAMP, B. H. (1938*a*). Notes on the distribution of the geometric mean. *Ann. Math. Stats.*, **9**, 221.

CAMP, B. H. (1938*b*). Further interpretations of the chi-square test. *J. Am. Stat. Ass.*, **33**, 537.

CAMPBELL, N. (1935). The statistical theory of errors. *Proc. Phys. Soc.*, **47**, 800.

CAMPBELL, N. (1939). Frequency interpretations in probability. *Nature*, **143**, 601.

CANNON, E. W., and WINTNER, A. (1935). An asymptotic formula for a class of distribution functions. *Proc. Edin. Math. Soc.*, **4**, 138.

CANTELLI, F. P. (1913). Sulla differenza media con ripetizione. *Giorn. Econ. e Riv. di Stat.*, February.

CANTELLI, F. P. (1916). La tendenza ad un limite nel senso del calcolo della probabilità. *Rend. Circ. Mat. di Palermo*, **16**, 191.

CANTELLI, F. P. (1917). Sulla probabilità come limite della frequenza. *Rend. R. Acc. Linc.* (5), **26**, 39.

CANTELLI, F. P. (1923). Sulla oscillazione delle frequenze intorno alla probabilità. *Metron*, **3**, No. 2, 167.

CANTELLI, F. P. (1929). Sulla legge di distribuzione dei redditi. *Giorn. Econ. e Riv. di Stat.*

CANTELLI, F. P. (1932). Una teoria astratta del calcolo della probabilità. *Giorn. Ist. Ital. Att.*, **3**, 257.

CANTELLI, F. P. (1933*a*). Considerazione sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del Sig. Paul Lévy. *Giorn. Ist. Ital. Att.*, **4**, 327.

CANTELLI, F. P. (1933*b*). Sulla determinazione empirica delle legge di probabilità. *Giorn. Ist. Ital. Att.*, **4**, 421.

CANTELLI, F. P. (1935). Considérations sur la convergence dans le calcul des probabilités. *Ann. Inst. H. Poincaré*, **5**, 1.

Cantelli, F. P. (1936). Considerazione su alcuni concette esposti nella introduzione della nota di R. de Mises. *Giorn. Ist. Ital. Att.*, **7**, 256.

Carleman, T. (1925). *Les fonctions quasi-analytiques.* Gauthier-Villars, Paris.

Carlson, J. L. (1932). A study of the distribution of means estimated from small samples by the method of maximum likelihood for Pearson's Type II curve. *Ann. Math. Stats.*, **3**, 86.

Carmichael, F. L. (1931). Methods of computing seasonal indices. *J. Am. Stat. Ass.*, **26**, 135.

Carslaw, H. S. (1930). *Introduction to the Theory of Fourier's Series and Integrals.* Macmillan, London.

Carver, H. C. (1932). Trapezoidal rule for computing seasonal indices. *Ann. Math. Stats.*, **3**, 361.

Carver, H. C. (1933). Note on the computation and modification of moments. *Ann. Math. Stats.*, **4**, 229.

Carver, H. C. (1936). The fundamental nature and proof of Sheppard's adjustments. *Ann. Math. Stats.*, **7**, 154.

Castellano, V. (1933*a*). Sulle relazioni tra curve di frequenza e curve di concentrazione e sui rapporti di concentrazione corrispondenti a determinate distribuzioni. *Metron*, **10**, No. 4, 3.

Castellano, V. (1933*b*). Sulla interpretazione dinamica del rapporto di concentrazione. *Giorn. Ist. Ital. Att.*, **4**, 268.

Castellano, V. (1934). Sulla scarto quadratico medio della probabilità di transvariazione. *Metron*, **11**, No. 4, 19.

Castellano, V. (1935). Recente letteratura sugli indici di variabilità. *Metron*, **12**, No. 3, 101.

Castellano, V. (1937). Sugli indici relativi di variabilità e sulla concentrazione dei caratteri con segno. *Metron*, **13**, No. 1, 31.

Castelnuovo, G. (1926–8). *Calcolo della probabilità.* Bologna.

Castelnuovo, G. (1932). Sur quelques problèmes se rattachant au calcul des probabilités. *Ann. Inst. H. Poincaré*, **3**, 465.

Cave, B. M., and Pearson, K. (1914). Numerical illustrations of the variate-difference correlation method. *Biom.*, **10**, 340.

Cave-Browne-Cave, F. E. (1904). On the influence of the time factor on the correlation between the barometric heights at stations more than 1000 miles apart. *Proc. Roy. Soc.*, **A**, **74**, 403.

Chandra Sekar, C., and Francis, M. G. (1941). A method to get the significance limit of a type of test criteria. *Sankhyā*, **5**, 165.

Chapelin, J. (1932). On a method of proceeding from partial cell-frequencies to ordinates and to total cell-frequencies in the case of a bivariate frequency surface. *Biom.*, **24**, 495.

Chapman, D. W. (1935). The generalised problem of correct matchings. *Ann. Math. Stats.*, **6**, 85.

Chapman, R. A. (1938). Applicability of the z-test to a Poisson distribution. *Biom.*, **30**, 188.

Charlier, C. V. L. (1906). Researches into the theory of probability. *Medd. Lunds Astr. Obs.*

Charlier, C. V. L. (1912). Contributions to the mathematical theory of statistics. *Medd. Lunds Astr. Obs.*

Charlier, C. V. L. (1928). A new form of the frequency function. *Medd. Lunds Astr. Obs.*, Series 2, No. 51.

Charlier, C. V. L. (1931). *Applications [de la théorie des probabilités] à l'astronomie.* (Part of the *Traité* edited by Borel.) Gauthier-Villars, Paris.

Cheshire, L., Oldis, E., and Pearson, E. S. (1932). Further experiments on the sampling distribution of the correlation coefficient. *J. Am. Stat. Ass.*, **27**, 121.

Chlodovsky, L. (1938). Le problème des moments et les polynomes de S. Bernstein. *Comptes rendus Acad. Sci. U.S.S.R.*, **19**, 659.

Christidis, B. G. (1931). The importance of the shape of plot in field experimentation. *J. Agr. Sci.*, **21**, 14.

CHURCH, A. E. R. (1925). On the moments of the distributions of squared standard deviations for samples of N drawn from an indefinitely large population. *Biom.*, **17**, 79.

CHURCH, A. E. R. (1926). On the means and squared standard deviations of small samples from any population. *Biom.*, **18**, 321.

CISBANI, R. (1938). Contributi alla teoria delle medie. *Metron*, **13**, No. 2, 23, and No. 3, 3.

CLAPHAM, A. R. (1931). Studies in sampling technique: cereal experiments. *J. Agr. Sci.*, **21**, 366 and 376.

CLAPHAM, A. R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *J. Ecology*, **24**, 232.

CLAREMONT, C. A. (1916). On the correlation between the 'corrected' cancer and diabetes death-rates. *Biom.*, **11**, 191.

CLARK, A., and LEONARD, W. H. (1939). The analysis of variance with special reference to data expressed as percentages. *J. Am. Soc. Agron.*, **31**, 55.

CLOPPER, C. J., and PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biom.*, **26**, 404.

COBB, C. W. (1939). Note on Frisch's diagonal regression. *Econometrika*, **7**, 77.

COCHRAN, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Camb. Phil. Soc.*, **30**, 178.

COCHRAN, W. G. (1935). A note on the influence of rainfall on the yield of cereals in relation to manurial treatment. *J. Agr. Sci.*, **25**, 510.

COCHRAN, W. G. (1936*a*). The χ^2-distribution for the binomial and Poisson series with small expectations. *Ann. Eug. Lond.*, **7**, 207.

COCHRAN, W. G. (1936*b*). Statistical analysis of field counts of diseased plants. *Supp. J.R.S.S.*, **3**, 49.

COCHRAN, W. G. (1937*a*). The efficiencies of the binomial series tests of significance of a mean and correlation coefficient. *J.R.S.S.*, **100**, 69.

COCHRAN, W. G. (1937*b*). Problems arising in the analysis of a series of similar experiments. *Supp. J.R.S.S.*, **4**, 102.

COCHRAN, W. G. (1938*a*). The omission or addition of an independent variate in multiple linear regression. *Supp. J.R.S.S.*, **5**, 171.

COCHRAN, W. G. (1938*b*). Some difficulties in the statistical analysis of replicated experiments. *Emp. J. Exp. Agr.*, **6**, 157.

COCHRAN, W. G. (1939*a*). Long-term agricultural experiments. *Supp. J.R.S.S.*, **6**, 104.

COCHRAN, W. G. (1939*b*). The use of the analysis of variance in enumeration by sampling. *J. Am. Stat. Ass.*, **34**, 492.

COCHRAN, W. G. (1940*a*). Note on an approximative formula for significance levels of z. *Ann. Math. Stats.*, **11**, 93.

COCHRAN, W. G. (1940*b*). The analysis of variance when experimental errors follow the Poisson or binomial laws. *Ann. Math. Stats.*, **11**, 335.

COCHRAN, W. G. (1941). The distribution of the largest of a set of variances as a fraction of their total. *Ann. Eug. Lond.*, **11**, 47.

COCHRAN, W. G. (1942*a*). The χ^2 correction for continuity. *Iowa State College J. Sci.*, **61**, 421.

COCHRAN, W. G. (1942*b*). Sampling theory when the sampling units are of unequal sizes. *J. Am. Stat. Ass.*, **37**, 199.

COCHRAN, W. G. (1943). The comparison of different scales of measurement for experimental results. *Ann. Math. Stats.*, **14**, 205.

COLEMAN, J. B. (1932). A coefficient of linear correlation based on the method of least squares and the line of best fit. *Ann. Math. Stats.*, **3**, 79.

COMRIE, L. J. (1936). Inverse interpolation and scientific applications of the National accounting machine. *Supp. J.R.S.S.*, **3**, 87.

COMRIE, L. J., HEY, G. B., and HUDSON, H. G. (1937). Application of Hollerith equipment to an agricultural investigation. *Supp. J.R.S.S.*, **4**, 210.

COMRIE, L. J. (1939). *Tables of $tan^{-1}\,x$ and $log\,(1 + x^2)$*. Tracts for Computers, No. 23. Cambridge University Press.

COMRIE, L. J., and HARTLEY, H. O. (1941). Tables of Lagrangian coefficients for harmonic interpolation in certain tables of percentage points. *Biom.*, **32**, 183.

Co-operative Study, see SOPER, H. E. and others, 1917.

COPELAND, A. H. (1928). Admissible numbers in the theory of probability. *Am. J. Maths.*, **50**, 535.

COPELAND, A. H. (1929). Independent event histories. *Am. J. Maths.*, **51**, 612.

COPELAND, A. H. (1932). The theory of probability from the point of view of admissible numbers. *Ann. Math. Stats.*, **3**, 143.

COPELAND, A. H. (1936). Point set theory applied to the random selection of the digits of an admissible number. *Am. J. Maths.*, **58**, 181.

COPELAND, A. H., and REGAN, F. (1936). A postulational treatment of the Poisson law. *Ann. Math.*, **37**, 357.

COPELAND, A. H. (1937). Consistency of conditions determining collectives. *Trans. Am. Math. Soc.*, **43**, 333.

CORNISH, E. A. (1936). Non-replicated factorial experiments. *J. Aus. Inst. Agr. Sci.*, **2**, 79.

CORNISH, E. A., and FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Stat.*, **5**, 307.

CORNISH, E. A. (1940*a*, *b*, *c*). The estimation of missing values in incomplete randomised block experiments. *Ann. Eug. Lond.*, **10**, 112; The estimation of missing values in quasi-factorial designs. *Ibid.*, **10**, 137; The analysis of covariance in quasi-factorial designs. *Ibid.*, **10**, 269.

COWLES, A. (1933). Can stock-market forecasters forecast? *Econometrika*, **1**, 309.

COWLES, A., and CHAPMAN, E. N. (1935). A statistical study of climate in relation to pulmonary tuberculosis. *J. Am. Stat. Ass.*, **30**, 517.

COWLES, A., and JONES, H. F. (1937). Some *a posteriori* probabilities in stock-market action. *Econometrika*, **5**, 280.

COX, G. M., and SNEDECOR, G. W. (1936). Covariance used to analyse the relation between corn-yield and average. *J. Farm. Econ.*, **18**, 597.

COX, G. M. (1940). Enumeration and construction of balanced incomplete block configurations. *Ann. Math. Stats.*, **11**, 72.

CRAIG, A. T. (1932). The simultaneous distribution of mean and standard deviation in small samples. *Ann. Math. Stats.*, **3**, 126.

CRAIG, A. T. (1933*a*). On the correlation between certain averages for small samples. *Ann. Math. Stats.*, **4**, 127.

CRAIG, A. T. (1933*b*). Variables correlated in sequence. *Bull. Am. Math. Soc.*, **39**, 129.

CRAIG, A. T. (1936*a*). Note on a certain bilinear form that occurs in statistics. *Am. J. Maths.*, **58**, 864.

CRAIG, A. T. (1936*b*). A certain mean-value problem in statistics. *Bull. Am. Math. Soc.*, **42**, 670.

CRAIG, A. T. (1938). On the independence of certain estimates of variance. *Ann. Math. Stats.*, **9**, 48.

CRAIG, A. T. (1939). On the mathematics of the representative method of sampling. *Ann. Math. Stats.*, **10**, 26.

CRAIG, A. T. (1943). Note on the independence of certain quadratic forms. *Ann. Math. Stats.*, **14**, 195.

CRAIG, C. C. (1928). An application of Thiele's seminvariants to the sampling problem. *Metron*, **7**, No. 4, 3.

CRAIG, C. C. (1929*a*). Sampling when the parent population is of Pearson's Type III. *Biom.*, **21**, 287.

CRAIG, C. C. (1929*b*). The frequency function of y/x. *Ann. Math.*, **30**, 471.

CRAIG, C. C. (1931*a*). Sampling in the case of correlated observations. *Ann. Math. Stats.*, **2**, 324.

CRAIG, C. C. (1931*b*). Note on the distribution of samples of N drawn from a Type A population. *Ann. Math. Stats.*, **2**, 99.

CRAIG, C. C. (1931*c*). On a property of the seminvariants of Thiele. *Ann. Math. Stats.*, **2**, 154.

CRAIG, C. C. (1932). On the composition of dependent elementary errors. *Ann. Math.*, **33**, 184.

CRAIG, C. C. (1933). On the Tchebycheff inequality of Bernstein. *Ann. Math. Stats.*, **4**, 94.

CRAIG, C. C. (1936*a*). On the frequency function of xy. *Ann. Math. Stats.*, **7**, 1.

CRAIG, C. C. (1936*b*). A new exposition and chart for the Pearson system of frequency curves. *Ann. Math. Stats.*, **7**, 16.

CRAIG, C. C. (1936*c*). Sheppard's corrections for a discrete variable. *Ann. Math. Stats.*, **7**, 55.

CRAIG, C. C. (1940). The product seminvariants of the mean and a central moment in samples. *Ann. Math. Stats.*, **11**, 177.

CRAIG, C. C. (1941*a*). Note on the distribution of non-central t with an application. *Ann. Math. Stats.*, **12**, 224.

CRAIG, C. C. (1941*b*). A note on Sheppard's corrections. *Ann. Math. Stats.*, **12**, 339.

CRAIG, J. I. (1916). A new method of discovering periodicities. *Month. Not. R. Astr. Soc.*, **76**, 493.

CRAMÉR, H. (1923). Das Gesetz von Gauss und die Theorie des Risikos. *Skand. Akt.*, **6**, 209.

CRAMÉR, H. (1926). On some classes of series used in mathematical statistics. *Skandinaviske Matematikercongres*, Copenhagen.

CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. Akt.*, **11**, 13 and 141.

CRAMÉR, H. (1934). Su un teorema relativo alla legge uniforme dei grandi numeri. *Giorn. Ist. Ital. Att.*, **5**, 1.

CRAMÉR, H. (1935*a*). Sur les propriétés asymptotiques d'une classe de variables aléatoires. *Comptes rendus*, **201**, 441.

CRAMÉR, H. (1935*b*). Sugli sviluppi asintotici di funzioni di repartizione in serie di polinomi di Hermite. *Giorn. Ist. Ital. Att.*, **6**, 141.

CRAMÉR, H. (1936). Über eine Eigenschaft der normalen Verteilungsfunktion. *Math. Zeit.*, **41**, 405.

CRAMÉR, H., and WOLD, H. (1936). Some theorems on distribution functions. *J. Lond. Math. Soc.*, **11**, 290.

CRAMÉR, H. (1937). *Random variables and probability distributions.* Cambridge University Press.

CRAMÉR, H. (1938–9). Entwicklungslinien der Wahrscheinlichkeitsrechnung. 9ᵉ *Congrès des Math. Scand.*, 67.

CRAMÉR, H., LÉVY, P., and VON MISES, R. (1938). Les sommes et les fonctions de variables aléatoires. *Conf. Int. de Sci. Math.*, **3**.

CROWTHER, G. (1934). The 'Economist' index of business activity. *J.R.S.S.*, **97**, 241.

CRUM, W. L. (1923). Cycles of rates on commercial paper. *Rev. Econ. Stats.*, **5**, 17.

CRUM, W. L. (1925). Progressive variation in seasonality. *J. Am. Stat. Ass.*, **20**, 48.

CRUM, W. L. (1933). An analytical interpretation of straw vote samples. *J. Am. Stat. Ass.*, **28**, 152.

CURETON, E. E., and DUNLAP, J. W. (1938). Developments in statistical methods related to test construction. *Rev. Educ. Res.*, **8**, 307.

CURTISS, J. H. (1941). On the distribution of the quotient of two chance variables. *Ann. Math. Stats.*, **12**, 409.

CURTISS, J. H. (1943). On transformations used in the analysis of variance. *Ann. Math. Stats.*, **14**, 107.

CZUBER, E. (1921). *Die statistische Forschungsmethode.* Seidel, Wien.

CZUBER, E. (1921, 1923). *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung.* Teubner, Leipzig.

DALY, J. F. (1940). On the unbiassed character of likelihood ratio tests for independence in normal systems. *Ann. Math. Stats.*, **11**, 1.

DANIELS, H. E. (1938*a*). The effect of departures from ideal conditions other than non-normality on the t- and z-tests of significance. *Proc. Camb. Phil. Soc.*, **34**, 321.

DANIELS, H. E. (1938*b*). Some problems of statistical interest in wool research. *Supp. J.R.S.S.*, **5**, 89.

DANIELS, H. E. (1941). A property of the distribution of extremes. *Biom.*, **32**, 194.

DANIELS, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biom.*, **33**, 129.

DANTZIG, G. B. (1939). On a class of distributions that approach the normal distribution function. *Ann. Math. Stats.*, **10**, 247.

DANTZIG, G. B. (1940). On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ. *Ann. Math. Stats.*, **11**, 186.

DARMOIS, G. (1928). *Statistique mathématique.* Octave Doin, Paris.

DARMOIS, G. (1929). Analyse et comparaison des séries statistiques qui se développent dans le temps. *Metron*, **8**, Nos. 1–2, 211.

DARMOIS, G. (1933). Distributions statistiques rattachées à la loi de Gauss et la répartition des revenus. *Econometrika*, **1**, 159.

DARMOIS, G. (1934). Sur la théorie des deux facteurs de Spearman. *Comptes rendus*, **199**, 1176 and 1358.

DARMOIS, G. (1935). Sur les lois de probabilité à estimation exhaustive. *Comptes rendus*, **200**, 1265.

DARMOIS, G. (1936). *L'emploi des observations statistiques. Méthodes d'estimation. Actualités scientifiques et industrielles*, No. 356. Paris. Hermann et Cie.

DAVID, F. N. (1934). On the $P_{\lambda n}$ test for randomness; remarks, further illustration and table for $P_{\lambda n}$. *Biom.*, **26**, 1.

DAVID, F. N. (1937). A note on unbiassed limits for the correlation coefficient. *Biom.*, **29**, 157.

DAVID, F. N. (1938*a*). *Tables of the Correlation Coefficient.* Cambridge University Press.

DAVID, F. N. (1938*b*). Limiting distributions connected with certain methods of sampling human populations. *Stat. Res. Mem.*, **2**, 69.

DAVID, F. N., and NEYMAN, J. (1938*c*). Extension of the Markoff theorem on least squares. *Stat. Res. Mem.*, **2**, 105.

DAVID, F. N. (1939). On Neyman's 'smooth' test for goodness of fit. I. Distribution of the criterion ψ^2 when the hypothesis tested is true. *Biom.*, **31**, 191.

DAVIES, G. R. (1930). First moment correlation. *J. Am. Stat. Ass.*, **25**, 413.

DAVIES, O. L. (1932). On the betas of quadrilateral distributions. *Biom.*, **24**, 498.

DAVIES, O. L. (1933, 1934). On asymptotic formulae for the hypergeometric series. I. Hypergeometric series in which the fourth element is unity. *Biom.*, **25**, 295; II. *Ibid.*, **26**, 59.

DAVIES, O. L., and PEARSON, E. S. (1934). Methods of estimating from samples the population standard deviation. *Supp. J.R.S.S.*, **1**, 76.

DAVIS, H. T. (editor) (1933, 1934). *Tables of the Higher Mathematical Functions. Parts I and II.* Bloomington Press, Indiana.

DAVIS, H. T. (1933). Polynomial approximation by the method of least squares. *Ann. Math. Stats.*, **4**, 155.

DAVIS, H. T. (1941). *The Analysis of Economic Time Series.* Bloomington Press, Indiana.

DAY, B., and FISHER, R. A. (1937). The comparison of variability in populations having unequal means. An example of the analysis of covariance with multiple dependent and independent variates. *Ann. Eug. Lond.*, **7**, 333.

DE FINETTI, B. (1929). Sulle funzioni a incremento aleatorio. *Rend. R. Acc. Linc.*, (6) **10**, 163.

DE FINETTI, B. (1930*a*). Le funzioni caratteristiche di legge istantanea. *Rend. R. Acc. Linc.*, (6) **12**, 278.

DE FINETTI, B., and PACIELLO, U. (1930*b*). Calcolo della differenza media. *Metron*, **8**, No. 3, 89.

DE FINETTI, B. (1931). Sui metodi proposti per il calcolo della differenza media. *Metron*, **9**, No. 1, 3.

DE FINETTI, B. (1932). Sulla legge di probabilità degli estremi. *Metron*, **9**, Nos. 3–4, 127.

DE FINETTI, B. (1933*a*). Classi di numeri aleatori equivalenti. *Rend. R. Acc. Linc.*, (6) **18**, 107; La legge dei grandi numeri nel caso dei numeri aleatori equivalenti. *Ibid.*, **18**, 203; and Sulla legge di distribuzione dei valori in una successione di numeri aleatori equivalenti. *Ibid.*, **18**, 279.

DE FINETTI, B. (1933*b*). Sull' approssimazione empirica di una legge di probabilità. *Giorn. Ist. Ital. Att.*, **4**, 415.

DE FINETTI, B. (1937). La prévision: ses logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, **7**, 1.

DE FINETTI, B. (1939*a*). Resoconto critico del colloquio di Ginevra intorno alla teoria delle probabilità. *Giorn. Ist. Ital. Att.*, **9**, 1.

DE FINETTI, B. (1939*b*). La teoria del rischio e il problema della rovina dei giocatori. *Giorn. Ist. Ital. Att.*, **10**, 41.

DEL CHIARO, A. (1936). Sui momenti delle leggi di distribuzione del Pólya a più variabili. *Giorn. Ist. Ital. Att.*, **7**, 151.

DELL' AGNOLA, C. A. (1937). Sulla tendenza ad una variabile casuale limite di una successione di variabili casuali punteggiate discontinue. *Att. Ist. Veneto Sci.*, **96**, 365.

DE LURY, D. (1938). Note on correlations. *Ann. Math. Stats.*, **9**, 149.

DEL VECCHIO, E. (1933). Sulla dipendenza statistica. *Giorn. Ist. Ital. Att.*, **4**, 235.

DEMING, W. E. (1931, 1934, 1935). On the application of least squares. I. *Phil. Mag.*, (7), **11**, 146; II. *Ibid.*, **17**, 804; III. *Ibid.*, **19**, 389.

DEMING, W. E. (1934, 1938). The chi-test and curve fitting. *J. Am. Stat. Ass.*, **29**, 372; and: Some thoughts on curve fitting and the chi-square test. *Ibid.*, **33**, 543.

DEMING, W. E., and BIRGE, R. T. (1934). On the statistical theory of errors. *Rev. Mod. Phys.*, **6**, No. 3, 122.

DEMING, W. E. (1937). On the significant figures of least squares and correlations. *Science*, **85**, 451.

DEMOIVRE, A. (1718). *The Doctrine of Chances.* (3rd edition, 1756.)

DENK, F. (1936). Über den Aufbau der Permutation geordnete Elemente. *J. für Math.*, **176**, 18.

DERKSON, J. B. D. (1939). On some infinite series introduced by Tschuprow. *Ann. Math. Stats.*, **10**, 380.

DETROIT EDISON CO. STATISTICAL DEPARTMENT (1930). A mathematical theory of seasonal indices. *Ann. Math. Stats.*, **1**, 57.

DE VERGOTTINI, M. (1936). *Relazioni fra gli indici di variabilità dei fenomeni collettivi composti e quelli dei fenomeni collettivi semplici.* Failli, Rome.

DIEULEFAIT, C. E. (1934*a*). Contribution à l'étude de la théorie de la correlation. *Biom.*, **26**, 379.

DIEULEFAIT, C. E. (1934*b*). Sur les développements des fonctions des fréquences en séries de fonctions orthogonales. *Metron*, **11**, No. 4, 77.

DIEULEFAIT, C. E. (1935*a*). Sur la corrélation au sens des modes. *Comptes rendus*, **200**, 1511.

DIEULEFAIT, C. E. (1935*b*). Généralisation des courbes de K. Pearson. *Metron*, **12**, No. 2, 95.

DIXON, W. J. (1940). A criterion for testing the hypothesis that two samples are from the same population. *Ann. Math. Stats.*, **11**, 199.

DIXON, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Stats.*, **15**, 119.

DODD, E. L. (1923). The greatest and the least variate under general laws of error. *Trans. Am. Math. Soc.*, **25**, 525.

DODD, E. L. (1926). The convergence of a general mean of measurements to the true value. *Bull. Am. Math. Soc.*, **32**, 282.

DODD, E. L. (1927). The convergence of general means and the invariance of form of certain frequency functions. *Am. J. Maths.*, **49**, 215.

DODD, E. L. (1930). The use of linear functions to detect hidden periods in data separated into small sets. *Ann. Math. Stats.*, **1**, 205.

DODD, E. L. (1931). Classification of sizes and measures by frequency functions. *J. Am. Stat. Ass.*, **26**, 277.

DODD, E. L. (1934). The complete independence of certain properties of means. *Ann. Math.*, **35**, 740.

DODD, E. L. (1937*a*). Internal and external means arising from the scaling of frequency functions. *Ann. Math. Stats.*, **8**, 12.

DODD, E. L. (1937*b*). Regression coefficients as means of certain ratios. *Am. Math. Monthly*, **44**, 306.

DODD, E. L. (1937*c*). Index numbers and regression coefficients as means, internal and external. *Rep. Third Ann. Res. Conf. Econ. Stat., Colorado Springs*, 13.

DODD, E. L. (1938). Interior and exterior means obtained by the method of moments. *Ann. Math. Stats.*, **9**, 153.

DODD, E. L. (1939*a*). The length of the cycles which result from the graduation of chance elements. *Ann. Math. Stats.*, **10**, 254.

DODD, E. L. (1939*b*). Periodogram analysis with the phase a chance variable. *Econometrika*, **7**, 57.

DODD, E. L. (1941*a*). The problem of assigning a length to the cycle to be found in a simple moving average and in a double moving average of chance data. *Econometrika*, **9**, 25.

DODD, E. L. (1941*b*). The cyclic effects of linear graduations persisting in the differences of the graduated values. *Ann. Math. Stats.*, **12**, 127.

DODD, E. L. (1942). Certain tests for randomness applied to data grouped into small sets. *Econometrika*, **10**, 249.

DODD, S. C. (1927). On criteria for factorising correlated variables. *Biom.*, **19**, 45.

DOEBLIN, W. (1936, 1937). Sur les chaînes discrètes de Markoff. *Comptes rendus*, **203**, 24 and 1210 ; and : Éléments d'une théorie générale des chaînes constantes simples de Markoff. *Ibid.*, **205**, 7.

DOEBLIN, W. (1937). Sur le cas continu des probabilités en chaîne. *Rend. R. Acc. Linc.*, **25**, 170 ; Le cas discontinu des probabilités en chaîne. *Pub. Fac. Sci. Univ. Masaryk*, No. 236, 3 ; and (with R. Fortet) : Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, **65**, 132.

DOEBLIN, W. (1938). Premiers éléments d'une étude systématique de l'ensemble de puissances d'une loi de probabilité. *Comptes rendus*, **206**, 306 ; and : Étude de l'ensemble de puissances d'une loi de probabilité. *Ibid.*, **206**, 718.

DOEBLIN, W. (1938, 1939). Sur les sommes d'un grand nombre de vecteurs aléatoires. *Comptes rendus*, **207**, 511 ; Sur certains mouvements aléatoires. *Ibid.*, **208**, 249 ; Sur les sommes d'un grand nombre de variables aléatoires indépendantes. *Bull. Sci. Math.*, (2), **63**, 23 and 35.

DOETSCH, G. (1934). Die in der Statistik seltener Ereignisse auftretenden Charlierschen Polynome und eine damit zusammenhängende Differenzialdifferenzgleichung. *Math. Ann.*, **109**, 257.

DONNER, O. (1928). *Die Saisonschwankungen als Problem der Konjunkturforschung.* Vierteljahrheften zur Konjunkturforschung, Sonderheft 6. Hobbing, Berlin.

DOOB, J. L. (1934*a*). Stochastic processes and statistics. *Proc. Nat. Acad. Sci.*, **20**, 376.

DOOB, J. L. (1934*b*). Probability and statistics. *Trans. Am. Math. Soc.*, **36**, 759.

DOOB, J. L. (1935). The limiting distributions of certain statistics. *Ann. Math. Stats.*, **6**, 160.

DOOB, J. L. (1936). Statistical estimation. *Trans. Am. Math. Soc.*, **39**, 410.

DOOB, J. L. (1937). Stochastic processes depending on a continuous parameter. *Trans. Am. Math. Soc.*, **42**, 107.

DOOB, J. L. (1938). Stochastic processes with an integral-valued parameter. *Trans. Am. Math. Soc.*, **44**, 87.

DOOB, J. L. (1941). Probability as measure. *Ann. Math. Stats.*, **12**, 206 (followed by discussion, DOOB and VON MISES, **12**, 215).

DOODSON, A. T. (1917). Relation of the Mode, Median and Mean in frequency curves. *Biom.*, **11**, 429.

DÖRGE, K. (1934). Eine Axiomatisierung der von Misesschen Wahrscheinlichkeitstheorie. *Jber. dtsch. Mat. Ver.*, **43**, 39.

DÖRGE, K. (1936). Zu der von R. v. Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. Zweite Mitteilung. Allgemeine Wahrscheinlichkeitstheorie. *Math. Zeit.*, **40**, 161.

DRESSEL, P. L. (1940). Statistical seminvariants and their estimates, with particular emphasis on their relation to algebraic invariants. *Ann. Math. Stats.*, **11**, 33.

DRESSEL, P. L. (1941). A symmetric method for obtaining unbiased estimates and expected values. *Ann. Math. Stats.*, **12**, 84.

DUBLIN, L. I., LOTKA, A. J., and SPIEGELMAN, M. (1935). The construction of life tables by correlation. *Metron*, **12**, No. 2, 121.

DUBOIS, P. (1939). Formulas and tables for rank correlation. *Psych. Rec.*, **3**, 46.

DUBORDIEU, J. (1939). *Théorie de l'assurance-maladie.* Paris.

DUGUÉ, D. (1936*a*). Sur le maximum de précision des estimations gaussiennes à la limite. *Comptes rendus*, **202**, 193 ; and : Sur le maximum de précision des lois limites d'estimation. *Ibid.*, **202**, 452.

DUGUÉ, D. (1936*b*). Sur certaines modes de convergence de lois d'estimation. *Comptes rendus*, **202**, 1732.

DUGUÉ, D. (1937*a*). Sur une extension de la loi des grands nombres. *Comptes rendus*, **204**, 317.

DUGUÉ, D. (1937*b*). Application des propriétés de la limite au sens du calcul des probabilités à l'étude des diverses questions d'estimation. *J. École Poly.*, **3**, No. 4, 305.

DUGUÉ, D. (1939). Sur quelques propriétés analytiques des fonctions caractéristiques. *Comptes rendus*, **208**, 1778.

DUNLAP, H. F. (1931). An empirical determination of the distribution of means, standard deviations and correlation coefficients drawn from rectangular populations. *Ann. Math. Stats.*, **2**, 66.

DWYER, P. S. (1937*a*). Moments of any rational integral isobaric sample moment function. *Ann. Math. Stats.*, **8**, 21.

DWYER, P. S. (1937*b*). The simultaneous computation of groups of regression equations and associated multiple regression coefficients. *Ann. Math. Stats.*, **8**, 224.

DWYER, P. S. (1938). Combined expansions of products of symmetric power sums and of sums of symmetric power products with application to sampling. *Ann. Math. Stats.*, **9**, 1 and 97.

DWYER, P. S. (1940). Combinatorial formulas for the rth standard moment of the sample sum, of the sample mean and of the normal curve. *Ann. Math. Stats.*, **11**, 353.

DWYER, P. S. (1941*a*). The solution of simultaneous equations. *Psychometrika*, **6**, 101.

DWYER, P. S. (1941*b*). The Doolittle technique. *Ann. Math. Stats.*, **12**, 449.

DWYER, P. S. (1941*c*). The skewness of the residuals in linear regression theory. *Ann. Math. Stats.*, **12**, 104.

DWYER, P. S. (1942). Recent developments in correlation technique. *J. Am. Stat. Ass.*, **37**, 441.

EDEN, T., and YATES, F. (1933). On the validity of Fisher's z-test when applied to an actual example of non-normal data. *J. Agr. Sci.*, **23**, 6.

EDGETT, G. L. (1931). Frequency distributions with given statistics which are not all moments. *Metron*, **9**, No. 2, 25.

EDGEWORTH, F. Y., generally ; see BOWLEY (1928).

EDGEWORTH, F. Y. (1905). The Law of Error. *Trans. Camb. Phil. Soc.*, **20**, 36 and 113 (with an Appendix not printed in the *T.C.P.S.* but issued with reprints).

EDGEWORTH, F. Y. (1906). The generalised law of error, or law of great numbers. *J.R.S.S.*, **69**, 497.

(GEWORTH, F. Y. (1908, 1909). On the probable errors of frequency constants. *J.R.S.S.*, **71**, 381, 499, 651, and **72**, 81.

(GEWORTH, F. Y. (1925*a*). Article 'Index numbers' in Palgrave's *Dictionary of Political Economy*, vol. 2, Macmillan.

(GEWORTH, F. Y. (1925*b*). The plurality of index-numbers. *Econ. J.*, **35**, 379.

(GEWORTH, F. Y. (1925*c*). The element of probability in index numbers. *J.R.S.S.*, **88**, 557.

LLS, W. C. (1929). Formulas for probable errors of coefficients of correlation. *J. Am. Stat. Ass.*, **24**, 170.

GENBERGER, F. (1924). Die Wahrscheinlichkeitsansteckung. *Mitt. Verein. Schweiz. Versich. Math.*, Heft **19**, 31.

SENHART, C. (1938). The power function of the χ^2-test. *Bull. Am. Math. Soc.*, **44**, 32.

SENHART, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Stats.*, **10**, 162.

DERTON, E. M. (1933). The Lanarkshire Milk Experiment. *Ann. Eug. Lond.*, **5**, 326.

DERTON, W. P. (1933). Adjustments for the moments of J-shaped curves. *Biom.*, **25**, 179.

DERTON, W. P., and HANSMANN, G. H. (1934). Improvement of curves fitted by the method of moments. *J.R.S.S.*, **97**, 330.

DERTON, SIR W. P. (1938*a*). *Frequency Curves and Correlation*, 3rd edn. Cambridge University Press.

DERTON, SIR W. P. (1938*b*). Correzioni dei momenti quando la curva è simmetrica. *Giorn. Ist. Ital. Att.*, **16**, 145.

FVING, G. (1937, 1938). Zur Theorie der Markoffschen Ketten. *Acta. Soc. Sci. Fennicae*, **2**, 1; and: Über die Interpretation von Markoffschen Ketten. *Soc. Sci. Fennicae Comment. phys.-nat.*, **10**, No. 3, 1.

SHANAWANY, M. R. (1936). An illustration of the accuracy of the χ^2-approximation. *Biom.*, **28**, 179.

IMETT, W. G. (1936). Sampling error and the two-factor theory. *Brit. J. Psych.*, **26**, 362.

IGELHART, M. D. (1936). The technique of path coefficients. *Psychometrika*, **1**, 287.

DÉLYI, A. (1937). Sulle connessioni fra due problemi di calcolo delle probabilità. *Giorn. Ist. Ital. Att.*, **8**, 328.

DÉLYI, A. (1938). Über eine erzeugende Funktion von Produkten Hermitescher Polynome. *Math. Zeit.*, **44**, 201.

DÖS, P., and TURAN, P. (1937, 1938). On Interpolation. I. Quadrature and mean-convergence in the Lagrange interpolation. *Ann. Math.*, **38**, 142; and II. On the distribution of fundamental points of Lagrange and Hermite interpolation. *Ibid.*, **39**, 703.

DÖS, P., and KAC, M. (1939). On the Gaussian law of errors in the theory of additive functions. *Proc. Nat. Acad. Sci.*, **25**, 206.

DÖS, P. (1939). On the smoothness of the asymptotic distribution of additive arithmetical functions. *Am. J. Math.*, **61**, 722.

DÖS, P., and WINTNER, A. (1939). Additive arithmetical functions and statistical independence. *Am. J. Maths.*, **61**, 713.

SCHER, F. (1932). On the probability function in the collective theory of risk. *Skand. Akt.*, **15**, 175.

ILER, L. (1782). Recherches sur une nouvelle espèce de quarrés magiques. *Verh. v. h. Zeeuwsch Genootsch. der Wetensch. Vlissingen*, 85.

RAUD, H. (1938*a*). Sur quelques lois d'erreurs à deux dimensions. *Comptes rendus*, **206**, 402.

RAUD, H. (1938*b*). Sur certaines décompositions en aléatoires imaginaires. *Comptes rendus*, **206**, 723.

SENCK, H. J. (1939). The validity of judgments as a function of the number of judges. *J. Exp. Psych.*, **25**, 650.

EKIEL, M. (1930*a*). *Methods of Correlation Analysis*. John Wiley and Sons, New York. (Chapman and Hall, London.)

EZEKIEL, M. (1930*b*). The sampling variability of linear and curvilinear regression. *Ann. Math. Stats.*, **1**, 275.

FALKNER, H. D. (1924). On the measurement of seasonal variations. *J. Am. Stat. Ass.*, **19**, 167.

FARR, W. (1919, 1920). Farr's law of density in relation to death rates. *J.R.S.S.*, **82**, 45, and **83**, 280.

FECHNER, G. T. (1897). *Kollektivmasslehre.* Engelmann, Leipzig.

FELD, W. (1924). Internationale Bibliographie der Statistik der Kindersterblichkeit. *Metron*, **3**, Nos. 3–4, 604.

FELDHEIM, E. (1936*a*). Sur l'orthogonalité des fonctions fondamentales de l'interpolation de Lagrange. *Comptes rendus*, **203**, 650.

FELDHEIM, E. (1936*b*). Sur les probabilités en chaîne. *Math. Ann.*, **112**, 775.

FELDHEIM, E. (1937*a*). Sulle legge di probabilità stabili a due variabili. *Giorn. Ist. Ital. Att.*, **8**, 146.

FELDHEIM, E. (1937*b*). Applicazioni dei polinomi di Hermite a qualche problema di calcolo delle probabilità. *Giorn. Ist. Ital. Att.*, **8**, 303.

FELDMAN, H. M. (1935). Mathematical expectation of product-moments of samples drawn from a set of infinite populations. *Ann. Math. Stats.*, **6**, 30.

FELLER, W. (1936*a*). Zur Theorie der stochastischer Prozesse. *Math. Ann.*, **113**, 113.

FELLER, W. (1936*b*, 1937). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Zeit.*, **40**, 521, and **42**, 301.

FELLER, W. (1937). Über das Gesetz der grossen Zahlen. *Acta Litt. Sci. Szeged*, **8**, 191.

FELLER, W. (1938). Note on regions similar to the sample space. *Stat. Res. Mem.*, **2**, 117.

FELLER, W. (1943). On a general class of 'contagious' distributions. *Ann. Math. Stats.*, **14**, 389.

FERTIG, J. W. (1936). On a method of testing the hypothesis that an observed sample of n variables and of size N has been drawn from a specified population of the same number of variables. *Ann. Math. Stats.*, **7**, 113.

FERTIG, J. W., and PROEHL, E. A. (1937). A test of a sample variance based on both tail ends of the distribution. *Ann. Math. Stats.*, **8**, 193.

FIELLER, E. C. (1931*a*). The duration of play. *Biom.*, **22**, 377.

FIELLER, E. C. (1931*b*). A problem in probability. *Biom.*, **22**, 425.

FIELLER, E. C. (1931*c*). The game of heads and tails. *Biom.*, **23**, 419.

FIELLER, E. C. (1932*a*). Numerical test of the adequacy of A. T. McKay's approximation. *J.R.S.S.*, **95**, 699.

FIELLER, E. C. (1932*b*). The distribution of an index in a normal bivariate population. *Biom.*, **24**, 428.

FIELLER, E. C. (1940). The biological standardisation of insulin. *Supp. J.R.S.S.*, **7**, 1.

FINNEY, D. J. (1938). The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population. *Biom.*, **30**, 190.

FINNEY, D. J. (1940, 1941, 1942). The detection of linkage. *Ann. Eug. Lond.*, **10**, 171 ; **11**, 10 ; **11**, 115 ; **12**, 31.

FINNEY, D. J. (1941*a*). The joint distribution of variance ratios based on a common-error mean square. *Ann. Eug. Lond.*, **11**, 136.

FINNEY, D. J. (1941*b*). On the distribution of a variate whose logarithm is normally distributed. *Supp. J.R.S.S.*, **7**, 155.

FISCHER, C. H. (1933*a*). On correlation surfaces of sums with a certain number of random elements in common. *Ann. Math. Stats.*, **4**, 103.

FISCHER, C. H. (1933*b*). On multiple and partial correlation coefficients of a certain sequence of sums. *Ann. Math. Stats.*, **4**, 278.

FISHER, A. (1922). *The Mathematical Theory of Probabilities and its application to Frequency-curves and Statistical Methods.* 2nd edn. Macmillan, New York.

FISHER, IRVING (1922). *The Making of Index Numbers.* Houghton Mifflin, Boston and New York.

FISHER, R. A. (1912). On an absolute criterion for fitting frequency curves. *Mess. Maths.*, **41**, 155.

FISHER, R. A. (1915). Frequency-distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biom.*, **10**, 507.

FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.*, **52**, 399.

FISHER, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. *Month. Not. R. Astr. Soc.*, **80**, 758.

FISHER, R. A. (1921*a*). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, **A**, **222**, 309.

FISHER, R. A. (1921*b*). Studies in crop-variation. I. An examination of the yield of dressed grain from Broadbalk. *J. Agr. Sci.*, **11**, 107.

FISHER, R. A. (1921*c*). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, No. 4, 1.

FISHER, R. A. (1922*a*). On the interpretation of χ^2 from contingency tables and the calculation of *P*. *J.R.S.S.*, **85**, 87.

FISHER, R. A. (1922*b*). The goodness of fit of regression formulae and the distribution of regression coefficients. *J.R.S.S.*, **85**, 597.

FISHER, R. A., THORNTON, H. G., and MACKENZIE, W. A. (1922*c*). The accuracy of the plating method of estimating the density of bacterial populations. *Ann. App. Biol.*, **9**, 325.

FISHER, R. A. (1923). Statistical tests of agreement between observation and hypothesis. *Economica*, **3**, 139.

FISHER, R. A. (1924*a*). The distribution of the partial correlation coefficient. *Metron*, **3**, 329.

FISHER, R. A. (1924*b*). The influence of rainfall on the yield of wheat at Rothamsted. *Phil. Trans. Roy. Soc.*, **B**, **213**, 89.

FISHER, R. A. (1924*c*). On a distribution yielding the error functions of several well-known statistics. *Proc. Int. Math. Congress*, Toronto, p. 805.

FISHER, R. A. (1924*d*). The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *J.R.S.S.*, **87**, 442.

FISHER, R. A. (1925*a*, 1944). *Statistical Methods for Research Workers.* (1st edn. 1925. 9th edn. 1944). Oliver and Boyd, Edinburgh.

FISHER, R. A. (1925*b*). Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700.

FISHER, R. A. (1926*a*). Applications of 'Student's' distribution. *Metron*, **5**, No. 3, 90; and: Expansion of 'Student's' integral in powers of n^{-1}. *Metron*, **5**, No. 3, 109.

FISHER, R. A. (1926*b*). On the random sequence. *Q.J. Roy. Met. Soc.*, **52**, 250.

FISHER, R. A. (1926*c*). Bayes' theorem and the fourfold table. *Eugenics Review*, **18**, 32.

FISHER, R. A., and WISHART, J. (1927). On the distribution of the error of an interpolated value and on the construction of tables. *Proc. Camb. Phil. Soc.*, **23**, 912.

FISHER, R. A., and TIPPETT, L. H. C. (1928*a*). Limiting forms of the frequency-distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, **24**, 180.

FISHER, R. A. (1928*b*). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.*, **A**, **121**, 654.

FISHER, R. A. (1928*c*). On a property connecting the χ^2 measure of discrepancy with the method of maximum likelihood. *Atti di Congresso Int. dei Matematici, Bologna*, **6**, 94.

FISHER, R. A. (1929*a*). Tests of significance in harmonic analysis. *Proc. Roy. Soc.*, **A**, **125**, 54.

FISHER, R. A. (1929*b*). Moments and product-moments of sampling distributions. *Proc. Lond. Math. Soc.*, (2), **30**, 199.

FISHER, R. A. (1930*a*). Inverse Probability. *Proc. Camb. Phil. Soc.*, **26**, 528.

FISHER, R. A. (1930*b*). The moments of the distribution for normal samples of measures of departure from normality. *Proc. Roy. Soc.*, **A**, **130**, 16.

FISHER, R. A., and WISHART, J. (1931). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. Lond. Math. Soc.*, **33**, 195.

FISHER, R. A. (1932). Inverse probability and the use of likelihood. *Proc. Camb. Phil. Soc.*, **28**, 257.

FISHER, R. A. (1933). The concepts of inverse probability and of fiducial probability referring to unknown parameters. *Proc. Roy. Soc.*, **A**, **139**, 343.

FISHER, R. A. (1934*a*). Two new properties of mathematical likelihood. *Proc. Roy. Soc.*, **A**, **144**, 285.

FISHER, R. A. (1934*b*). Probability, likelihood and quantity of information in the logic of uncertain inference. *Proc. Roy. Soc.*, **A**, **146**, 1.

FISHER, R. A., and YATES, F. (1934*c*). The 6 × 6 Latin square. *Proc. Camb. Phil. Soc.*, **30**, 492.

FISHER, R. A. (1934*d*). The effect of methods of ascertainment upon the estimation of frequencies. *Ann. Eug. Lond.*, **6**, 13.

FISHER, R. A. (1935*a*). The logic of inductive inference. *J.R.S.S.*, **98**, 39.

FISHER, R. A. (1935*b*). The fiducial argument in statistical inference. *Ann. Eug. Lond.*, **6**, 391.

FISHER, R. A. (1935*c*, 1942). *The Design of Experiments* (1st edn. 1935, 3rd edn. 1942). Oliver and Boyd, Edinburgh.

FISHER, R. A. (1936*a*). The use of multiple measurements in taxonomic problems. *Ann. Eug. Lond.*, **7**, 179.

FISHER, R. A. (1936*b*). The coefficient of racial likeness. *J. Roy. Anthrop. Soc.*, **66**, 57.

FISHER, R. A. (1936*c*). Uncertain inference. *Proc. Roy. Soc.*, **B**, **122**, 1.

FISHER, R. A. (1937*a*). Professor Karl Pearson and the method of moments. *Ann. Eug. Lond.*, **7**, 303.

FISHER, R. A. (1937*b*). On a point raised by M. S. Bartlett on fiducial probability. *Ann. Eug. Lond.*, **7**, 370.

FISHER, R. A., and YATES, F. (1938*a*, 1942). *Statistical Tables for use in Biological, Agricultural and Medical Research.* 2nd edn. 1942. Oliver and Boyd, Edinburgh.

FISHER, R. A. (1938*b*). Quelques remarques sur l'estimation statistique. *Biotypologie*, **6**, 153.

FISHER, R. A. (1938*c*). The statistical utilisation of multiple measurements. *Ann. Eug. Lond.*, **8**, 376.

FISHER, R. A. (1938*d*). *Statistical Theory of Estimation.* Calcutta Readership Lectures. Published by the University of Calcutta.

FISHER, R. A. (1939*a*). The comparison of samples with possibly unequal variances. *Ann. Eug. Lond.*, **9**, 174.

FISHER, R. A. (1939*b*). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eug. Lond.*, **9**, 238.

FISHER, R. A. (1940*a*). On the similarity of the distributions found for the test of significance in harmonic analysis and in Stevens's problem in geometrical probability. *Ann. Eug. Lond.*, **10**, 14.

FISHER, R. A. (1940*b*). An examination of the different possible solutions of a problem in incomplete blocks. *Ann. Eug. Lond.*, **10**, 52.

FISHER, R. A. (1940*c*). A note on fiducial inference. *Ann. Math. Stats.*, **10**, 383.

FISHER, R. A. (1940*d*). The precision of discriminant functions. *Ann. Eug. Lond.*, **10**, 422.

FISHER, R. A. (1941*a*). The asymptotic approach to Behrens' integral with further tables for the *d*-test of significance. *Ann. Eug. Lond.*, **11**, 141.

FISHER, R. A. (1941*b*). The negative binomial distribution. *Ann. Eug. Lond.*, **11**, 182.

FISHER, R. A. (1942*a*). New cyclic solutions to problems in incomplete blocks. *Ann. Eug. Lond.*, **11**, 290.

FISHER, R. A. (1942*b*). The likelihood solution of a problem in compounded probabilities. *Ann. Eug. Lond.*, **11**, 306.

FISHER, R. A. (1942*c*). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eug. Lond.*, **11**, 341.

FISHER, R. A. (1942*d*). Some combinatorial theorems and enumerations connected with the numbers of diagonal types in a Latin square. *Ann. Eug. Lond.*, **11**, 395.

FISHER, R. A. (1942*e*). Completely orthogonal 9 × 9 squares : a correction. *Ann. Eug. Lond.*, **11**, 402.

FLUX, A. W. (1921, 1933). The measurement of price changes. *J. Roy. Stat. Soc.*, **84**, 167, and **96**, 606.

FORTET, R. (1935–8). Sur les probabilités en chaîne. *Comptes rendus*, **201**, 184, **202**, 1362, and **204**, 315 ; and : Sur l'itération des substitutions algébriques linéaires à une infinité de variables et ses applications à la théorie des probabilités en chaîne. *Rev. Ci., Lima*, **40**, 185, 337, 481.

FRANKEL, A., and KULLBACK, S. (1940). A simple sampling experiment on confidence intervals. *Ann. Math. Stats.*, **11**, 209.

FRANKEL, L. R., and HOTELLING, H. (1938). The transformation of statistics to simplify their distribution. *Ann. Math. Stats.*, **9**, 87.

FRANKEL, L. R., and STOCK, J. S. (1939). The allocation of samplings among several strata. *Ann. Math. Stats.*, **10**, 288.

FRÉCHET, M. (1930). Sur la convergence en probabilité. *Metron*, **8**, No. 4, 3.

FRÉCHET, M., and SHOHAT, J. (1931). A proof of the generalised second-limit theorem. *Trans. Am. Math. Soc.*, **33**, 533.

FRÉCHET, M. (1933). Sur le coefficient, dit de corrélation, et sur la corrélation en général. *Rev. Inst. Int. Stat.*, **4**, 1.

FRÉCHET, M. (1935). Sur l'équation fonctionnelle de Chapman et sur le problème des probabilités en chaîne. *Proc. Lond. Math. Soc.*, **39**, 515.

FRÉCHET, M. (1936*a*). Sull' espressione esatta di uno scarto medio. *Giorn. Ist. Ital. Att.*, **6**, 164.

FRÉCHET, M. (1936*b*). Sul caso positivamente regolare nel problema delle probabilità concatenate. *Giorn. Ist. Ital. Att.*, **7**, 28.

FRÉCHET, M. (1937*a*). Sulla mescolanza delle palline e sulle leggi-limite delle probabilità. *Giorn. Ist. Ital. Att.*, **8**, 14.

FRÉCHET, M. (1937*b*). *Recherches théoriques modernes.* (Part of the *Traité* edited by Borel.) Gauthier-Villars, Paris.

FRICKEY, E. (1937). The theory of index-number bias. *Rev. Econ. Stat.*, **19**, 161.

FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Ass.*, **32**, 675.

FRIEDMAN, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *Ann. Math. Stats.*, **11**, 86.

FRISCH, R. (1926). Sur les semi-invariants et moments employés dans l'étude des distributions statistiques. *Oslo, Skrifter af det Norske Videnskaps Academie, II, Hist.-Filos. Klasse*, No. 3.

FRISCH, R. (1928). Changing harmonics and other general types of components in empirical series. *Skand. Akt.*, **11**, 220.

FRISCH, R. (1929). Correlation and scatter in statistical variables. *Nordisk. Stat. J.*, **1**, 36.

FRISCH, R. (1930). Necessary and sufficient conditions regarding the form of an index-number which shall meet certain of Fisher's tests. *J. Am. Stat. Ass.*, **25**, 397.

FRISCH, R. (1931). A method of decomposing an empirical series into its cyclical and progressive components. *J. Am. Stat. Ass.*, **26**, Supp. p. 73.

FRISCH, R., and MUDGETT, B. D. (1931). Statistical correlation and the theory of cluster types. *J. Am. Stat. Ass.*, **26**, 375.

FRISCH, R. (1932). On the use of difference equations in the study of frequency-distributions. *Metron*, **10**, No. 3, 35.

FRISCH, R. (1933). Propagation problems and impulse problems in dynamic economics. *Economic Essays in honour of Gustav Cassel.* London.

FRISCH, R. (1934*a*). Robert Schmidt's definition of skewness and kurtosis. *Econometrika*, **2**, 221.

FRISCH, R. (1934*b*). Statistical confluence analysis by means of complete regression equations. Publication No. 5, Universitets Ökonomiske Institut, Oslo.

FRISCH, R. (1936). Annual survey of general economic theory. The problem of index numbers. *Econometrika*, **4**, 1.

FRISCH, R. (1938). On the inversion of moving averages. *Skand. Akt.*, **21**, 218.

FRY, T. C. (1928). *Probability and its Engineering Uses.* van Nostrand, New York.

FRY, T. C. (1938). The χ^2-test of significance. *J. Am. Stat. Ass.*, **33**, 513.

GALTON, SIR FRANCIS (1886). Regression towards mediocrity in hereditary stature. *J. Anthrop. Inst.*, **15**, 246 ; and : Family likeness in stature. *Proc. Roy. Soc.*, **A**, **40**, 42.

GALTON, SIR FRANCIS (1902). The most suitable proportion between the values of first and second prizes. *Biom.*, **1**, 385.

GALVANI, L. (1931). Contributi allà determinazione degli indici di variabilità per alcuni tipi di distribuzione. *Metron*, **9**, No. 1, 3.

GALVANI, L. (1932). Sulle curve di concentrazione relative a caratteri limitate e non limitate. *Metron*, **10**, No. 3, 61.

GARNER, R. (1932). Concerning the limits of a measure of skewness. *Ann. Math. Stats.*, **3**, 358.

GARWOOD, F. (1933). The probability integral of the correlation coefficient in samples from a normal bivariate population. *Biom.*, **25**, 71.

GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biom.*, **28**, 437.

GARWOOD, F. (1940). An application of the theory of probability to the operation of vehicular-controlled traffic signals. *Supp. J.R.S.S.*, **7**, 65.

GARWOOD, F. (1941). The application of maximum likelihood to dosage-mortality curves. *Biom.*, **32**, 46.

GEARY, R. C. (1927). Some properties of correlation and regression in a limited universe. *Metron*, **7**, No. 1, 83.

GEARY, R. C. (1930). The frequency distribution of the quotient of two normal variables. *J.R.S.S.*, **93**, 442.

GEARY, R. C. (1933). A general expression for the moments of certain symmetrical functions of normal samples. *Biom.*, **25**, 184.

GEARY, R. C. (1935*a*). The ratio of the mean deviation to the standard deviation as a test of normality. *Biom.*, **27**, 310.

GEARY, R. C. (1935*b*). Note on the correlation between β_2 and w'. *Biom.*, **27**, 353.

GEARY, R. C. (1936*a*). Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biom.*, **28**, 295.

GEARY, R. C. (1936*b*). The distribution of 'Student's' ratio for non-normal samples. *Supp. J.R.S.S.*, **3**, 178.

GEARY, R. C., and PEARSON, E. S. (1938). *Tests of Normality.* Biometrika Office, London.

GEARY, R. C. (1942). The estimation of many parameters. *J.R.S.S.*, **105**, 213.

GEARY, R. C. (1943). Minimum range for quasi-normal distributions. *Biom.*, **33**, 100.

GEARY, R. C. (1944). Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter. *Biom.*, **33**, 123.

GEHLKE, C. E., and BIEHL, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J. Am. Stat. Ass.*, **29**, *Supp.*, 169.

GEIRINGER, H. (1933). Korrelationsmessung auf Grund der Summenfunktion. *Zeit. ang. Math. und Mech.*, **13**, 121.

GEIRINGER, H. (1934). Une méthode générale de statistique théorique. *Comptes rendus*, **198**, 420 ; and : Applications. *Ibid.*, **198**, 696.

GEIRINGER, H. (1938). On the probability theory of arbitrarily linked events. *Ann. Math. Stats.*, **9**, 260 (and *Errata*, **10**, 202).

GEIRINGER, H. (1942). A new explanation of non-normal dispersion in the Lexis theory. *Econometrika*, **10**, 53.

GINI, C. (1912). Variabilità e Mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche. Studi Economico-Giuridici delle R. Università di Cagliari.

GINI, C. (1916). Indici di concordanza. *Atti R. Ist. Veneto di Sci. Lett. ed Arte.*

GINI, C. (1921). Sull' interpolazione di una retta quando i valori della variabile indipendente sono affetti da errori accidentali. *Metron*, **1**, No. 3, 63.

GINI, C., and GALVANI, L. (1929). Di talune estensioni del concetto di media ai caratteri qualitativi. *Metron*, **8**, Nos. 1–2, 3.

GINI, C. (1930). Sul massimo degli indici di variabilità assoluta, etc. *Metron*, **8**, No. 3, 3.

GINI, C. (1932). Intorno alle curve di concentrazione. *Metron*, **9**, Nos. 3–4, 3.

GINI, C. (1938). Di una formola comprensiva delle medie. *Metron*, **13**, No. 2, 3.

GINI, C., and ZAPPA, G. (1938). Sulle proprietà delle medie potenziate e combinatorie. *Metron*, **13**, No. 3, 21.

GINI, C. (1939). Sulla determinazione dell' indice di cograduazione. *Metron*, **13**, No. 4, 41.

GIRSHIK, M. A. (1936). Principal components. *J. Am. Stat. Ass.*, **31**, 519.

GIRSHIK, M. A. (1939). On the sampling theory of the roots of determinantal equations. *Ann. Math. Stats.*, **10**, 203.

GIRSHIK, M. A. (1942). Note on the distribution of roots of a polynomial with random complex coefficients. *Ann. Math. Stats.*, **13**, 235; Correction, *ibid.*, **13**, 447.

GLIVENKO, V. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Att.*, **4**, 92.

GLIVENKO, V. (1936). Sul teorema limite della teoria delle funzioni caratteristiche. *Giorn. Ist. Ital. Att.*, **7**, 160.

GNEDENKO, B. (1938). Über die Konvergenz der Verteilungsgesetze von Summen voneinander unabhängiger Summanden. *C.R. Acad. Sci. U.S.S.R.*, **18**, 231.

GONIN, H. T. (1936). The use of factorial moments in the treatment of the hypergeometric distribution and in tests for regression. *Phil. Mag.*, (7), **21**, 215.

GORDON, R. A. (1937). A selected bibliography of the literature of economic fluctuations. *Rev. Econ. Stat.*, **19**, 37.

GORDON, R. D. (1939). Estimating bacterial populations by the dilution method. *Biom.*, **31**, 167.

GORDON, R. D. (1941). The estimation of a quotient when the denominator is normally distributed. *Ann. Math. Stats.*, **12**, 115.

GOTAAS, P. (1936). Formules de récurrence pour les semi-invariants à quelques lois de distribution à plusieurs variables. *Comptes rendus*, **202**, 619.

GOULDEN, C. H. (1937). Efficiency in field trials of pseudo-factorial and incomplete randomised block methods. *Canadian J. Res.*, **15**, 231.

GOULDEN, C. H. (1938). Modern methods for testing a large number of varieties. *Dom. Canada Dep. Agr. Tech. Bull.*, 9.

GOULDEN, C. H. (1939). *Methods of Statistical Analysis.* John Wiley and Sons, New York. (Chapman and Hall, London.)

GRAM, J. P. (1879). *Om Rackkendviklinger bestemte ved Hjaelp af de mindste Kvadraters Methode*, Copenhagen. Reprinted as Über die Entwicklung realer Funktionen in Reihen mittelst der Methode der kleinsten Quadraten. *J. für Math.*, **94**, 41, 1894.

GREENLEAF, H. E. H. (1932). Curve approximation by means of functions analogous to the Hermite polynomials. *Ann. Math. Stats.*, **3**, 204.

GREENSTEIN, B. (1935). Periodogram analysis with special application to business failures in the United States. *Econometrika*, **3**, 170.

GREENWOOD, J. A., and STUART, C. E. (1937). Mathematical techniques used in extra-sensory perception research. *J. Parapsychology*, **1**, 206.

GREENWOOD, J. A. (1938). Variance of a general matching problem. *Ann. Math. Stats.*, **9**, 56.

Greenwood, J. A., and Greville, T. N. E. (1939). On the probability of attaining a given standard deviation ratio in an infinite series of trials. *Ann. Math. Stats.*, **10**, 297.

Greenwood, J. A. (1940). The first four moments of a general matching problem. *Ann. Eug. Lond.*, **10**, 290.

Greenwood, M., and Yule, G. U. (1915). The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. *Proc. Roy. Soc. Medicine*, **8**, 113.

Greenwood, M., and Yule, G. U. (1917). On the statistical interpretation of some bacteriological methods employed in water analysis. *J. Hygiene*, **21**, 36.

Greenwood, M., and Yule, G. U. (1920). An inquiry into the nature of frequency-distributions of multiple happenings, etc. *J.R.S.S.*, **83**, 255.

Greenwood, M. (1922). The value of life tables in statistical research. *J.R.S.S.*, **85**, 537.

Gressens, O. (1925). On the measurement of seasonal variations. *J. Am. Stat. Ass.*, **20**, 203.

Greville, T. N. E. (1938). Exact probabilities for the matching hypothesis. *J. Parapsychology*, **2**, 55.

Greville, T. N. E. (1939). Invariance of the admissibility of numbers under certain general types of transformations. *Trans. Am. Math. Soc.*, **46**, 410.

Greville, T. N. E. (1941). The frequency-distribution of a general matching problem. *Ann. Math. Stats.*, **12**, 350.

Grüneberg, H., and Haldane, J. B. S. (1937). Tests of goodness of fit applied to records of Mendelian segregation in mice. *Biom.*, **29**, 144.

Guldberg, A. (1922). Sur un théorème de M. Markoff. *Comptes rendus*, **175**, 679.

Guldberg, A. (1934). On discontinuous frequency functions of two variables. *Skand. Akt.*, **17**, 89.

Guldberg, A. (1935). Sur les lois de probabilités et la corrélation. *Ann. Inst. H. Poincaré*, **5**, 159.

Guldberg, S. (1935). Sui momenti della legge di distribuzioni del Pólya. *Giorn. Ist. Ital. Att.*, **6**, 394.

Gulotta, B. (1938). Sulla legge di probabilità della differenza tra la media empirica e il valore medio teorico dei quadrati d' una variabile casuale che segue la legge normale. *Giorn. Ist. Ital. Att.*, **9**, 245.

Gumbel, E. J. (1924). Eine Darstellung der Sterbetafel. *Biom.*, **16**, 283 (and Correction, *ibid.*, 411).

Gumbel, E. J. (1925). Lebenserwartung und mittleres Alter der Lebenden. *Biom.*, **17**, 173.

Gumbel, E. J. (1932). La distribuzione dei decessi secondo la legge di Gauss. *Giorn. Ist. Ital. Att.*, **3**, 311.

Gumbel, E. J. (1934). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré*, **5**, 115.

Gumbel, E. J. (1935*a*). Les *m*-ièmes valeurs extrêmes et le logarithme du nombre d'observations. *Comptes rendus*, **200**, 509.

Gumbel, E. J. (1935*b*). Le plus grand âge, distribution et série. *Comptes rendus*, **201**, 318.

Gumbel, E. J. (1937). La durée extrême de la vie humaine. *Actualités Scientifiques et Industrielles*, No. 520. Paris. Hermann et Cie.

Gumbel, E. J. (1938*a*). La prévision des inondations. *Comptes rendus*, **206**, 558; and: La distribution des inondations, *Akt. Vedy Roc.*, **7**, 85.

Gumbel, E. J. (1938*b*). Gli eventi compatibili. *Giorn. Ist. Ital. Att.*, **9**, 3 and 58.

Gumbel, E. J. (1939). Les valeurs de position d'une variable aléatoire. *Comptes rendus*, **208**, 149.

Gumbel, E. J. (1941). The return period of flood flows. *Ann. Math. Stats.*, **12**, 163.

Gumbel, E. J. (1942). Simple tests for given hypotheses. *Biom.*, **32**, 317.

Gumbel, E. J. (1943*a*). On serial numbers. *Ann. Math. Stats.*, **14**, 163.

Gumbel, E. J. (1943*b*). On the reliability of the classical χ^2-test. *Ann. Math. Stats.*, **14**, 253.

HAAVELMO, T. (1941). A note on the variate-difference method. *Econometrika*, **9**, 74.

HABERLER, G. (1927). Der Sinn der Indexzahlen. Mohr, Tübingen.

HADAMARD, J., and FRÉCHET, M. (1933). Sur les probabilités discontinues des événements en chaîne. *Zeit. ang. Math. und Mech.*, **13**, 92.

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small. *Biom.*, **29**, 133.

HALDANE, J. B. S. (1938, 1939, 1940). The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small. *Biom.*, **29**, 389; The mean and variance of χ^2 when used as a test of homogeneity when samples are small. *Biom.*, **31**, 346; The cumulants and moments of the binomial distribution and the cumulants of χ^2 for an $n \times 2$-fold table. *Biom.*, **31**, 392; Corrections to formulae in papers on the moments of χ^2. *Biom.*, **31**, 220.

HALDANE, J. B. S. (1938). The approximate normalisation of a class of frequency-distributions. *Biom.*, **29**, 392.

HALDANE, J. B. S. (1941). The cumulants of the distribution of the square of a variate. *Biom.*, **32**, 199.

HALDANE, J. B. S. (1942*a*). Moments of the distributions of powers and products of normal variates. *Biom.*, **32**, 226.

HALDANE, J. B. S. (1942*b*). The mode and median of a nearly normal distribution with given cumulants. *Biom.*, **32**, 294.

HALL, P. (1927*a*). Multiple and partial correlation coefficients in the case of an n-fold variate system. *Biom.*, **19**, 100.

HALL, P. (1927*b*). The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biom.*, **19**, 240.

HALPHEN, E. (1939). Sur la convergence des estimations. *Comptes rendus*, **208**, 708.

HAMBURGER, H. (1920, 1921). Über eine Erweiterung des Stieltjesschen Momentproblems. *Math. Ann.*, **81**, 235; **82**, 120 and 168.

HANSEN, M. H., and HURWITZ, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Stats.*, **14**, 333.

HANSMANN, G. H. (1934). On certain non-normal symmetrical frequency-distributions. *Biom.*, **26**, 129.

HARRIS, J. A. (1914). On the calculation of intra-class and inter-class coefficients of correlation from class-moments when the number of possible combinations is large. *Biom.*, **9**, 446.

HARRIS, J. A., and GUNSTAD, B. (1931). Extension of Pearson's correlation method to intra-class and inter-class relationships. *J. Agr. Sci.*, **42**, 279.

HARRIS, J. A., and TRELOAR, A. E. (1927). On a limitation in the applicability of the contingency coefficient. *J. Am. Stat. Ass.*, **22**, 460; and: HARRIS and CHI TU. A second category of limitations in the applicability of the contingency coefficient. *Ibid.*, **24**, 367. (Reply by K. Pearson, *J. Am. Stat. Ass.*, **25**, 320.)

HART, B. I. (1942). Significance levels for the ratio of the mean square successive difference to the variance. *Ann. Math. Stats.*, **13**, 445.

HARTLEY, H. O. (1938). Studentisation and large sample theory. *Supp. J.R.S.S.*, **5**, 80.

HARTLEY, H. O. (1940). Testing the homogeneity of a set of variances. *Biom.*, **31**, 249.

HARTLEY, H. O. (1942). The range in normal samples. *Biom.*, **32**, 334.

HARTLEY, H. O. (1944). Studentization, or the elimination of the standard deviation of the parent population from the random sample-distribution of statistics. *Biom.*, **33**, 173.

HARTMAN, P., VAN KAMPEN, E. R., and WINTNER, A. (1937). Mean motions and distribution functions. *Am. J. Maths.*, **59**, 261.

HARTMAN, P., VAN KAMPEN, E. R., and WINTNER, A. (1938). On the distribution functions of almost periodic functions. *Am. J. Maths.*, **60**, 491.

HARTMAN, P., VAN KAMPEN, E. R., and WINTNER, A. (1939). Asymptotic distributions and statistical independence. *Am. J. Maths.*, **61**, 477.

HARZER, P. (1933). *Tabellen für alle Statistischen Zwecke.* Abhandlungen des Bayerischen Ak. der Wiss., Math-naturwiss. Abteilung, Neue Folge, Heft 21.

HAUSSDORF, F. (1923). Momentprobleme für ein endliches Intervall. *Math. Zeit.*, **16**, 220.

HAVILAND, E. K. (1934*a*). On the theory of absolutely additive distribution functions. *Am. J. Maths.*, **56**, 625.

HAVILAND, E. K. (1934*b*). On distribution functions and their Laplace-Fourier transform. *Proc. Nat. Acad. Sci.*, **20**, 50 ; and (with A. WINTNER) : On the Fourier-Stieltjes transform. *Am. J. Maths.*, **56**, 1.

HAVILAND, E. K. (1935). On the inversion formula for Fourier-Stieltjes transforms in more than one dimension. *Am. J. Maths.*, **57**, 94, and **57**, 382. Also : Note, **57**, 569.

HAVILAND, E. K. (1935, 1936). On the moment problem for distribution functions in more than one dimension. *Am. J. Maths.*, **57**, 562, and **58**, 164.

HAVILAND, E. K. (1939). Asymptotic probability distributions and harmonic curves. *Am. J. Maths.*, **61**, 947.

HELGUERO, F. (1906). Per la risoluzione delle curve dimorfiche. *Rend. R. Acad. Linc.*, **6**.

HELMERT, F. R. (1875). Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler. *Zeit. für Math. und Phys.*, **20**, 300.

HELMERT, F. R. (1876*a*). Über die Wahrscheinlichkeit der Potenzsummen und über einige damit in Zusammenhange stehende Fragen. *Zeit. für Math. und Phys.*, **21**, 192.

HELMERT, F. R. (1876*b*). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers direkter Beobachtungen gleicher Genauigkeit. *Astronomische Nachrichten*, **88**, No. 2096.

HENDERSON, J. (1922). On expansions in tetrachoric functions. *Biom.*, **14**, 157.

HENDERSON, R. (1907). Frequency curves and moments. *J. Inst. Act.*, **41**, 429.

HENDRICKS, W. A. (1931). The use of the relative residual in the application of the method of least squares. *Ann. Math. Stats.*, **2**, 458.

HENDRICKS, W. A. (1934). The standard error of any analytic function of a set of parameters evaluated by the method of least squares. *Ann. Math. Stats.*, **5**, 107.

HENDRICKS, W. A. (1935). The analysis of variance considered as an application of simple error theory. *Ann. Math. Stats.*, **6**, 117.

HENDRICKS, W. A. (1936). An approximation to 'Student's' distribution. *Ann. Math. Stats.*, **7**, 210.

HENDRICKS, W. A., and ROBEY, K. W. (1936). The sampling distribution of the coefficient of variation. *Ann. Math. Stats.*, **7**, 129.

HERSCH, L. (1934). Essai sur les variations périodiques et leur mensuration. *Metron*, **12**, No. 1, 3.

HEY, G. B. (1938). A new method of experimental sampling illustrated on certain non-normal populations. *Biom.*, **30**, 68.

HILDEBRANDT, E. H. (1931). Systems of polynomials connected with the Charlier expansions and the Pearson differential equations. *Ann. Math. Stats.*, **2**, 379.

HILTON, J. (1924, 1928). Enquiry by sample ; an experiment and its results. *J.R.S.S.*, **87**, 544 ; and : Some further enquiries by sample. *Ibid.*, **91**, 519.

HIRSCHFELD, H. O. (1935). A connection between correlation and contingency. *Proc. Camb. Phil. Soc.*, **31**, 520.

HIRSCHFELD, H. O. (1937). The distribution of the ratio of covariance estimates in two samples drawn from normal bivariate populations. *Biom.*, **29**, 65.

HOEL, P. G. (1937). A significance test for component analysis. *Ann. Math. Stats.*, **8**, 149.

HOEL, P. G. (1938). On the chi-square distribution for small samples. *Ann. Math. Stats.*, **9**, 158.

HOEL, P. G. (1939). A significance test for minimum rank in factor analysis. *Psychometrika*, **4**, 245.

HOEL, P. G. (1941). On methods of solving normal equations. *Ann. Math. Stats.*, **12**, 354.

HOJO, T. (1931, 1933). Distribution of the median, quartiles and interquartile distance in samples from a normal population. *Biom.*, **23**, 315 ; and : A further note on the relation between the median and the quartiles in small samples from a normal population. *Biom.*, **25**, 79.

HOLZINGER, K. J., and CHURCH, A. E. R. (1929). On the means of samples from a U-shaped population. *Biom.*, **20A**, 361.

HORST, P. (1935). A method for determining the coefficients of a characteristic equation. *Ann. Math. Stats.*, **6**, 83.

HOSTINSKY, B. (1937). Sur les probabilités relatives aux variables aléatoires liées entre elles. Applications diverses. *Ann. Inst. H. Poincaré*, **7**, 69.

HOTELLING, H. (1925). The distribution of correlation ratios calculated from random data. *Proc. Nat. Acad. Sci.*, **11**, 657.

HOTELLING, H. (1927). An application of analysis situs to statistics. *Bull. Am. Math. Soc.*, July-Aug., 467.

HOTELLING, H. (1930). The consistency and ultimate distribution of optimum statistics. *Trans. Am. Math. Soc.*, **32**, 847.

HOTELLING, H. (1931). The generalisation of 'Student's' ratio. *Ann. Math. Stats.*, **2**, 360.

HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. Reprinted from *J. Educ. Psych.* (**24**, 417), Sept.-Oct. 1933, Warwick and York, Inc., Baltimore.

HOTELLING, H. (1936*a*). Simplified calculation of principal components. *Psychometrika*, **1**, 27.

HOTELLING, H. (1936*b*). Relations between two sets of variates. *Biom.*, **28**, 321.

HOTELLING, H., and PABST, M. R. (1936*c*). Rank correlation and tests of significance involving no assumptions of normality. *Ann. Math. Stats.*, **7**, 29.

HOTELLING, H., and FRANKEL, L. R. (1938). The transformation of statistics to simplify their distribution. *Ann. Math. Stats.*, **9**, 87.

HOTELLING, H. (1939). Tubes and spheres in n-spaces and a class of statistical problems. *Am. J. Maths.*, **61**, 440.

HOTELLING, H. (1940). The selection of variates for use in prediction with some comments on the problem of nuisance parameters. *Ann. Math. Stats.*, **11**, 271.

HOTELLING, H. (1941). Experimental determination of the maximum of a function. *Ann. Math. Stats.*, **12**, 20.

HOTELLING, H. (1943). Some new methods in matrix calculation. *Ann. Math. Stats.*, **14**, 1 and 440.

HSU, C. T., and LAWLEY, D. N. (1939). The derivation of the fifth and sixth moments of b_2 in samples from a normal population. *Biom.*, **31**, 238.

HSU, C. T. (1940, 1941). On samples from a normal bivariate population. *Ann. Math. Stats.*, **11**, 410 ; and : Samples from two bivariate normal populations. *Ibid.*, **12**, 279.

HSU, P. L. (1938*a*). Contribution to the theory of 'Student's' t-test as applied to the problem of two samples. *Stat. Res. Mem.*, **2**, 1.

HSU, P. L. (1938*b*). On the best unbiassed quadratic estimate of the variance. *Stat. Res. Mem.*, **2**, 91.

HSU, P. L. (1938*c*). Notes on Hotelling's generalised T. *Ann. Math. Stats.*, **9**, 231.

HSU, P. L. (1939*a*). A new proof of the joint product-moment distribution. *Proc. Camb. Phil. Soc.*, **35**, 336.

HSU, P. L. (1939*b*). On the distribution of roots of certain determinantal equations. *Ann. Eug. Lond.*, **9**, 250.

HSU, P. L. (1940). On generalised analysis of variance. *Biom.*, **31**, 221.

HSU, P. L. (1941*a*). On the limiting distribution of the canonical correlations. *Biom.*, **32**, 38.

HSU, P. L. (1941*b*). Analysis of variance from the power function standpoint. *Biom.*, **32**, 62.

HSU, P. L. (1941*c*). On the problem of rank and the limiting distribution of Fisher's test function. *Ann. Eug. Lond.*, **11**, 39.

Hsu, P. L. (1941*d*). Canonical reduction of the general regression problem. *Ann. Eug. Lond.*, **11**, 42.

Hsu, P. L. (1943). Some simple facts about the separation of degrees of freedom in factorial experiments. *Sankhyā*, **6**, 253.

Immer, F. R. (1937). Correlation between means and standard deviations in field experiments. *J. Am. Stat. Ass.*, **32**, 525.

Ingham, A. E. (1933). An integral which occurs in statistics. *Proc. Camb. Phil. Soc.*, **29**, 270.

Irwin, J. O. (1925*a*). The further theory of Francis Galton's individual difference problem. *Biom.*, **17**, 100.

Irwin, J. O. (1925*b*). On a criterion for the rejection of outlying observations. *Biom.*, **17**, 238.

Irwin, J. O. (1927, 1929). On the frequency-distribution of the means of samples from a population having any law of frequency with finite moments, etc. *Biom.*, **19**, 225, and **21**, 431.

Irwin, J. O. (1929*a*). On the frequency-distribution of any number of deviates from the mean of a sample from a normal population and the partial correlations between them. *J.R.S.S.*, **92**, 580.

Irwin, J. O. (1929*b*). Note on the χ^2-test for goodness of fit. *J.R.S.S.*, **92**, 274.

Irwin, J. O. (1930). On the frequency-distribution of the means of samples from populations of certain of Pearson's types. *Metron*, **8**, No. 4, 51.

Irwin, J. O. (1931). Mathematical theorems involved in the analysis of variance. *J.R.S.S.*, **94**, 284.

Irwin, J. O. (1933). A critical discussion of the single-factor theory. *Brit. J. Psych.*, **23**, 371.

Irwin, J. O. (1934). On the independence of the constituent items in the analysis of variance. *Supp. J.R.S.S.*, **1**, 236.

Irwin, J. O. (1935). Tests of significance for differences between percentages based on small numbers. *Metron*, **12**, No. 2, 83.

Irwin, J. O. (1937*a*). The frequency-distribution of the difference between two independent variates following the same Poisson distribution. *J.R.S.S.*, **100**, 415.

Irwin, J. O. (1937*b*). Statistical method applied to biological assays. *Supp. J.R.S.S.*, **4**, 1.

Irwin, J. O., and Cheeseman, E. A. (1939). On the maximum-likelihood method of determining dosage response curves. *Supp. J.R.S.S.*, **6**, 174.

Irwin, J. O. (1942). On the distribution of a weighted estimate of variance and on analysis of variance in certain cases of unequal weighting. *J.R.S.S.*, **105**, 115.

Irwin, J. O., and Kendall, M. G. (1944). Sampling moments of moments for a finite population. *Ann. Eug. Lond.*, **12**, 138.

Isserlis, L. (1914, 1916). On the partial correlation ratio. Part I. Theoretical. *Biom.*, **10**, 391; and Part II. Numerical. *Ibid.*, **11**, 50.

Isserlis, L. (1915). On the conditions under which the probable errors of frequency-distributions have a real significance. *Proc. Roy. Soc.*, **A**, **92**, 23. (Correction, *Biom.*, **12**, 261.)

Isserlis, L. (1916). On certain probable errors and correlation coefficients of multiple frequency-distributions with skew regression. *Biom.*, **11**, 185.

Isserlis, L. (1917). On the representation of statistical data. *Biom.*, **11**, 418.

Isserlis, L. (1918*a*). On the value of a mean as calculated from a sample. *J.R.S.S.*, **81**, 75.

Isserlis, L. (1918*b*). On a formula for the product-moment coefficient of any order of a normal frequency-distribution in any number of variables. *Biom.*, **12**, 134. (Correction, *ibid.*, **12**, 266.)

Isserlis, L. (1918*c*). Formulae for determining the mean values of products of deviations of mixed moment coefficients in two to eight variables in samples taken from a limited population. *Biom.*, **12**, 183.

Isserlis, L. (1931). On the moment distributions of moments in the case of samples drawn from a limited universe. *Proc. Roy. Soc.*, **A**, **132**, 586.

Isserlis, L. (1936). Inverse probability. *J.R.S.S.*, **99**, 130.

JACKSON, D. (1921). Note on the median of a set of numbers. *Bull. Am. Math. Soc.*, **27**, 160.

JACKSON, D. (1934). Series of orthogonal polynomials. *Ann. Math.*, **34**, 527; and: The summation of series of orthogonal polynomials. *Bull. Am. Math. Soc.*, **40**, 743.

JACKSON, D. (1937). Orthogonal polynomials on a plane curve. *Duke Math. J.*, **3**, 228.

JACKSON, D. (1938). Orthogonal polynomials in three variables. *Duke Math. J.*, **4**, 441.

JACKSON, R. W. (1936). Tests of statistical hypotheses in the case when the set of alternatives is discontinuous, illustrated on some genetical problems. *Stat. Res. Mem.*, **1**, 138.

JACOB, M. (1933). Sullo sviluppo di una curva di frequenza in serie di Charlier Type B. *Giorn. Ist. Ital. Att.*, **4**, 221.

JACOB, M. (1935, 1937). Sul fenomeno di Gibbs nello sviluppo in serie di polinomi di Hermite. *Giorn. Ist. Ital. Att.*, **6**, 1, and **8**, 297.

JEFFREYS, H. (1933). On Gauss's proof of the law of errors. *Proc. Camb. Phil. Soc.*, **29**, 231.

JEFFREYS, H. (1937*a*). On statistically steady distributions in Astronomy. *Monthly Not. R. Astr. Soc.*, **97**, 59.

JEFFREYS, H. (1937*b*). On the relation between direct and inverse methods in statistics. *Proc. Roy. Soc.*, **A**, **160**, 325.

JEFFREYS, H. (1937*c*). The law of errors and the combination of observations. *Phil. Trans.*, **A**, **237**, 231.

JEFFREYS, H. (1938*a*). Significance tests for continuous departures from suggested distributions of chance. *Proc. Roy. Soc.*, **A**, **164**, 307.

JEFFREYS, H. (1938*b*). The use of minimum χ^2 as an approximation to the method of maximum likelihood. *Proc. Camb. Phil. Soc.*, **34**, 156.

JEFFREYS, H. (1938*c*). Maximum likelihood, inverse probability and the method of moments. *Ann. Eug. Lond.*, **8**, 146.

JEFFREYS, H. (1938*d*). The correction of frequencies for a known standard error of observations. *Month. Not. R. Astr. Soc.*, **98**, 190.

JEFFREYS, H. (1939*a*). *The Theory of Probability.* Cambridge University Press.

JEFFREYS, H. (1939*b*). The minimum χ^2 approximation. *Proc. Camb. Phil. Soc.*, **35**, 520.

JEFFREYS, H. (1939*c*). The posterior probability distributions of the ordinary and intra-class correlation coefficients. *Proc. Roy. Soc.*, **A**, **167**, 464.

JEFFREYS, H. (1939*d*). The comparison of series of measures on different hypotheses concerning the standard errors. *Proc. Roy. Soc.*, **A**, **167**, 367.

JEFFREYS, H. (1939*e*). Random and systematic arrangements. *Biom.*, **31**, 1.

JEFFREYS, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eug. Lond.*, **10**, 48.

JEFFREYS, H. (1941). Some applications of the method of minimum χ^2. *Ann. Eug. Lond.*, **11**, 108.

JENKINS, T. N. (1932). A short method and tables for the calculation of the average and standard deviation of logarithmic distributions. *Ann. Math. Stats.*, **3**, 45.

JENNETT, W. J., and WELCH, B. L. (1939). The control of proportion defective as judged by a single quality characteristic varying on a continuous scale. *Supp. J.R.S.S.*, **6**, 80.

JENSEN, A. (1925). Report on the representative method in statistics. *Bull. Int. Stat. Inst.*, **22**, I^er livre.

JESSEN, B., and WINTNER, A. (1935). Distribution functions and the Riemann zeta-function. *Trans. Am. Math. Soc.*, **38**, 48.

JOHNSON, E. (1940). Estimates of parameters by means of least squares. *Ann. Math. Stats.*, **11**, 453.

JOHNSON, N. L., and WELCH, B. L. (1939). On the calculation of the cumulants of the χ-distribution. *Biom.*, **31**, 216.

JOHNSON, N. L., and WELCH, B. L. (1940*a*). Applications of the non-central t-distribution. *Biom.*, **31**, 362.

JOHNSON, N. L. (1940*b*). Parabolic test for linkage. *Ann. Math. Stats.*, **11**, 227.
JOHNSON, P. O., and NEYMAN, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Stat. Res. Mem.*, **1**, 57.
JONES, H. E. (1937*a*). Some geometrical considerations in the general theory of fitting lines and planes. *Metron*, **13**, No. 1, 21.
JONES, H. E. (1937*b*). The nature of regression functions in the correlation analysis of time-series. *Econometrika*, **5**, 305.
JONES, H. E. (1937*c*). The theory of runs as applied to time series. Report, Third Annual Research Conf. on Economics and Statistics, p. 33. (Cowles Commission.)
JORDAN, C. (1927). *Statistique Mathématique.* Gauthier-Villars, Paris.
JORDAN, C. (1932). Approximation and graduation according to the principle of least squares by orthogonal polynomials. *Ann. Math. Stats.*, **3**, 257.
JORDAN, C. (1933). Inversione della formola di Bernouilli relativa al problema delle prove ripetute a più variabili. *Giorn. Ist. Ital. Att.*, **4**, 505.
JORDAN, C. (1934). Teoria della perequazione e dell' approssimazione. *Giorn. Ist. Ital. Att.*, **5**, 81.
JÖRGENSEN, N. R. (1916). *Undersögelser over Frequensflader og Korrelation.* Busck, Copenhagen.

KAC, M. (1939). On a characterisation of the normal distribution. *Am. J. Maths.*, **61**, 726.
KAC, M., and VAN KAMPEN, E. R. (1939). Circular equidistributions and statistical independence. *Am. J. Maths.*, **61**, 677.
KALECKI, M. (1935). A macrodynamic theory of business cycles. *Econometrika*, **3**, 327.
KAMKE, E. (1932). *Einführung in die Wahrscheinlichkeitstheorie.* Hirzel, Leipzig.
KAPLANSKY, J. (1939). On a generalisation of the ' problème de rencontres '. *Am. Math. Monthly*, **46**, 159.
KAPTEYN, J. C. (1903). *Skew Frequency-Curves in Biology and Statistics.* Noordhoff, Groningen and Wm. Dawson, London.
KAUCKY, J. (1936). Le problème des itérations dans un cas de probabilités dépendantes. *Comptes rendus*, **202**, 722.
KELLEY, T. L. (1923). *Statistical Method.* Macmillan, New York.
KELLEY, T. L. (1928). *Cross-roads in the Mind of Man.* Stanford University Press, California.
KELLEY, T. L., and MCNEMAR, Q. (1929). Doolittle versus Kelley–Salisbury iteration method for computing multiple regression coefficients. *J. Am. Stat. Ass.*, **24**, 164.
KELLEY, T. L. (1935). An unbiassed correlation ratio measure. *Proc. Nat. Acad. Sci.*, **21**, 554.
KELLEY, T. L. (1938). *The Kelley Statistical Tables.* Macmillan, New York.
KENDALL, M. G. (1938*a*). The conditions under which Sheppard's corrections are valid. *J.R.S.S.*, **101**, 592.
KENDALL, M. G. (1938*b*). A new measure of rank correlation. *Biom.*, **30**, 81.
KENDALL, M. G., KENDALL, S. F. H., and BABINGTON SMITH, B. (1939). The distribution of Spearman's coefficient of rank correlation, etc. *Biom.*, **30**, 251.
KENDALL, M. G., and BABINGTON SMITH, B. (1939*a*). *Tables of Random Sampling Numbers.* Tracts for Computers, No. 24, Cambridge University Press.
KENDALL, M. G., and BABINGTON SMITH, B. (1939*b*). The problem of m rankings. *Ann. Math. Stats.*, **10**, 275.
KENDALL, M. G., and BABINGTON SMITH, B. (1940). On the method of paired comparisons. *Biom.*, **31**, 324.
KENDALL, M. G. (1940). Some properties of k-statistics. *Ann. Eug. Lond.*, **10**, 106 ; Proof of Fisher's rules for ascertaining the sampling semi-invariants of k-statistics. *Ibid.*, **10**, 215 ; The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. *Ibid.*, **10**, 392.
KENDALL, M. G. (1941). A theory of randomness. *Biom.*, **32**, 1.
KENDALL, M. G. (1942*a*). Partial rank correlation. *Biom.*, **32**, 277.

KENDALL, M. G. (1942*b*). On seminvariant statistics. *Ann. Eug. Lond.*, **11**, 300.

KENDALL, M. G. (1944*a*). Oscillatory movements in English agriculture. *J.R.S.S.*, **106**, 91.

KENDALL, M. G. (1944*b*). On autoregressive time-series. *Biom.*, **33**, 105.

KERCHNER, R., and WINTNER, A. (1936). On the asymptotic distribution of almost periodic functions with linearly independent frequencies. *Am. J. Maths.*, **58**, 91.

KERMACK, W. O., and MCKENDRICK, A. G. (1936). Tests for randomness in a series of numerical observations. *Proc. Roy. Soc. Edin.*, **57**, 228.

KERMACK, W. O., and MCKENDRICK, A. G. (1937). Some distributions associated with a randomly arranged set of numbers. *Proc. Roy. Soc. Edin.*, **57**, 332.

KERRICH, J. E. (1935). Systems of osculating arcs. *J. Inst. Act.*, **66**, 88.

KERRICH, J. E. (1937). Least squares and a generalisation of the 'Student'-Fisher theorem. *Skand. Akt.*, **20**, 244.

KEYFITZ, N. (1938). Graduation by a truncated normal. *Ann. Math. Stats.*, **9**, 66.

KEYNES, J. M. (1911). Principal averages and the laws of error which lead to them. *J.R.S.S.*, **74**, 322.

KEYNES, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

KHINTCHINE, A. (1928). Begründung der Normalkorrelation nach der Lindebergschen Methode. *Nachr. Forschungsinst. Moskau*, **1**.

KHINTCHINE, A. (1932–1933). Sulle successione stazionarie di eventi. *Giorn. Ital. Ist. Att.*, **3**, 267; and Über stationäre Reihen zufälliger Variablen. *Rec. Mathématiques, Moscou*, **40**.

KHINTCHINE, A. (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Springer, Berlin.

KHINTCHINE, A. (1934). Korrelationstheorie der stationäre stochastischer Prozesse. *Math. Ann.*, **109**, 604.

KHINTCHINE, A. (1935). Sul dominio di attrazione della legge di Gauss. *Giorn. Ist. Ital. Att.*, **6**, 378.

KHINTCHINE, A., and LÉVY, P. (1936). Sur les lois stables. *Comptes rendus*, **202**, 374.

KHINTCHINE, A. (1937*a*). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Rec. Math. Moscou*, **2**, 79.

KHINTCHINE, A. (1937*b*). Series of papers on probability laws in *Bull. Univ. État Moscou, Sér. Int. Sect. A*, **1**, Fasc. 1, 1, 6; Fasc. 5, 1, 4, 6.

KHINTCHINE, A. (1938). Zwei Sätze über stochastische Prozesse mit stabilen Verteilungen. *Rec. Math. Moscou*, **3**, 577.

KIBBLE, W. F. (1941). A two-variate gamma type distribution. *Sankhyā*, **5**, 137.

KISER, C. V. (1934). Pitfalls in sampling for population study. *J. Am. Stat. Ass.*, **29**, 250.

KISHEN, K. (1940). On a simplified method of expressing the components of the second-order interaction in a 3^3 factorial design. *Sankhyā*, **4**, 577.

KISHEN, K. (1942). Symmetrical unequal block arrangements. *Sankhyā*, **5**, 329.

KITAGAWA, T. (1941). The limit theorems of the stochastic contagious processes. *Mem. Fac. Sci.*, Kyusyu Imperial University, **A, 1**, 167.

KOLMOGOROFF, A. (1929). Bemerkungen zu meiner Arbeit 'Über die Summen zufälliger Grössen.' *Math. Ann.*, **102**, 484.

KOLMOGOROFF, A. (1931). Uber die analytische Methode in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, **104**, 415.

KOLMOGOROFF, A. (1933*a*). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin.

KOLMOGOROFF, A. (1933*b*). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Att.*, **4**, 83.

KOLMOGOROFF, A. (1937*a*). Zur Umkehrbarkeit der statistischen Naturgesetz. *Math. Ann.*, **113**, 766.

KOLMOGOROFF, A. (1937*b*). Chaînes de Markoff avec une infinité dénombrable des états possibles. *Bull. Univ. État Moscou, Sér. Int. Sect. A*, **1**, Fasc. 3, 1.

KOLMOGOROFF, A. (1941). Confidence limits for an unknown distribution function. *Ann. Math. Stats.*, **12**, 461.

KOLODZIECZYK, ST. (1933). Sur l'erreur de la seconde catégorie dans le problème de M. Student. *Comptes rendus*, **197**, 814.

KOLODZIECZYK, ST. (1935). On an important class of statistical hypothesis. *Biom.*, **27**, 161.

KONDO, T. (1929). On the standard error of the mean square contingency. *Biom.*, **21**, 376.

KONDO, T. (1930). A theory of the sampling distribution of standard deviations. *Biom.*, **22**, 36.

KONÖS, A. A. (1939). The problem of the true index number of the cost of living. *Econometrika*, **7**, 10.

KOOPMAN, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Am. Math. Soc.*, **39**, 399.

KOOPMANS, T. (1937). Linear regression analysis of economic time series. *Neth. Econ. Inst.*, No. 20. Haarlem.

KOOPMANS, T. (1940). The degree of damping in business cycles. *Econometrika*, **8**, 79.

KOOPMANS, T. (1941). Distributed lags in dynamic economics. *Econometrika*, **9**, 128.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Stats.*, **13**, 14.

KOSHAL, R. S. (1933). Application of the method of maximum likelihood in the improvement of curves fitted by the method of moments. *J.R.S.S.*, **96**, 303.

KOSHAL, R. S. (1935). Application of the method of maximum likelihood to the derivation of efficient statistics for fitting frequency curves. *J.R.S.S.*, **98**, 128.

KOSHAL, R. S. (1939). Maximal likelihood and minimal χ^2 in relation to frequency curves. *Ann. Eug. Lond.*, **9**, 209.

KOZAKIEWICZ, M. W. (1937, 1938). Sur les conditions nécéssaires et suffisantes pour la convergence stochastique. *Comptes rendus*, **205**, 1028 and *Fund. Math.*, **31**, 160.

KULLBACK, S. (1934). An application of characteristic functions to the distribution problem of statistics. *Ann. Math. Stats.*, **5**, 264.

KULLBACK, S. (1935*a*). On samples from a multivariate normal population. *Ann. Math. Stats.*, **6**, 202.

KULLBACK, S. (1935*b*). On the Bernouilli distribution. *Bull. Am. Math. Soc.*, **41**, 857.

KULLBACK, S. (1935*c*). A note on the distribution of a certain partial belonging coefficient. *Metron*, **12**, No. 3, 65.

KULLBACK, S. (1936*a*). The distribution laws of the difference and quotient of variables independently distributed in Pearson Type III laws. *Ann. Math. Stats.*, **7**, 51.

KULLBACK, S. (1936*b*). On certain distribution theorems of statistics. *Bull. Am. Math. Soc.*, **42**, 407.

KULLBACK, S. (1936*c*). A note on the multiple correlation coefficient. *Metron*, **12**, No. 4, 67.

KULLBACK, S. (1937). On certain distributions derived from the multinomial distribution. *Ann. Math. Stats.*, **8**, 127.

KUNETZ, G. (1936). Sur quelques propriétés des fonctions caractéristiques. *Comptes rendus*, **202**, 1829.

KUZMIN, R. O. (1939). Sur la loi de distribution du coefficient de corrélation dans les tirages d'un ensemble normal. *C.R. Acad. Sci. U.S.S.R.*, **22**, 298.

KUZNETS, S. (1929). Random events and cyclical fluctuations. *J. Am. Stat. Ass.*, **24**, 258.

KUZNETS, S. (1933). *Seasonal Patterns in Industry and Trade.* New York.

LADERMAN, J. (1939). The distribution of 'Student's' ratio for sampling of two items drawn from non-normal universes. *Ann. Math. Stats.*, **10**, 376.

LADERMAN, J., and LOWAN, A. N. (1939). On the distribution of *n*th tabular differences. *Ann. Math. Stats.*, **10**, 360.

LANDAHL, H. D. (1938). Centroid orthogonal transformations. *Psychometrika*, **3**, 219.

LAPLACE, P. S., MARQUIS DE (1818). *Théorie analytique des probabilités.*

LARMOR, SIR J., and YAMAGA, N. (1917). On permanent periodicity in sunspots. *Proc. Roy. Soc.*, A, **93**, 493.

LÁSKA, V. (1935). Contribution à la standardisation des définitions des principales notions statistiques. *Rev. Stat. Tchéchoslovaque*, **16**, 3.

LAWLEY, D. N. (1938). A generalisation of Fisher's z-test. *Biom.*, **30**, 180; and Correction, *ibid.*, **30**, 467.

LE CORBEILLER, P. (1933). Les systèmes auto-entretenus et les oscillations de rélaxation. *Econometrika*, **1**, 328.

LEDERMANN, W. (1938). The orthogonal transformation of a factorial matrix into itself. *Psychometrika*, **3**, 181.

LEDERMANN, W. (1939). Sampling distribution and selection in a normal population. *Biom.*, **30**, 295.

LEHMANN, A. (1939). Über die Inversion des Gausschen Wahrscheinlichkeits-Integrals. *Mitt. Verein. Schweiz. Versicherungs-Math.*, **38**, 15.

LENGYEL B. A. (1939). On testing the hypothesis that two samples have been drawn from a common normal population. *Ann. Math. Stats.*, **10**, 365.

LE ROUX, J. M. (1931). A study of the distribution of variance in small samples. *Biom.* **23**, 134.

LESER, C. E. V. (1942). Inequalities for multivariate frequency-distributions. *Biom.* **32**, 284.

LÉVY, P. (1925). *Calcul des Probabilités.* Gauthier-Villars, Paris.

LÉVY, P. (1931*a*). Quelques théorèmes sur les probabilités dénombrables. *Comptes rendus*, **192**, 658.

LÉVY, P. (1931*b*). Sulla legge forte dei grandi numeri. *Giorn. Ist. Ital. Att.*, **4**, 1.

LÉVY, P. (1931*c*). Sur un théorème de M. Khintchine. *Bull. Sci. Math.*, (2), **55**, 145.

LÉVY, P. (1934). Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali R. Sci. Norm. Sup. Pisa*, (2), **3**, 337.

LÉVY, P. (1935*a*). Sull' applicazione della geometria dello spazio di Hilbert allo studio delle successioni di variabili casuali. *Giorn. Inst. Ital. Att.*, **6**, 13.

LÉVY, P. (1935*b*). Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées. *J. Math. Pur. App.*, (7), **14**, 347.

LÉVY, P. (1936*a*). Sur quelques points de la théorie des probabilités dénombrables. *Ann. Inst. H. Poincaré*, **6**, 153.

LÉVY, P. (1936*b*). Détermination générale des lois limites. *Comptes rendus*, **203**, 698.

LÉVY, P. (1936*c*). La loi forte des grands nombres pour les variables aléatoires enchaînées. *J. Math. Pur. App.*, **15**, 11.

LÉVY, P. (1937*a*). L'arithmétique des lois de probabilité. *Comptes rendus*, **204**, 80 and 944; Sur les exponentielles des polynômes et sur l'arithmétique des produits de lois de Poisson. *Ann. l'École Norm. Sup.*, **54**, 231; and: Nouvelle contribution à l'arithmétique des produits de lois de Poisson. *Comptes rendus*, **205**, 535.

LÉVY, P. (1938*a*). L'arithmétique des lois de probabilité. *J. Math. Pur. App.*, **27**, 17.

LÉVY, P. (1938*b*, 1939). Sur la définition des lois de probabilités par leurs projections. *Comptes rendus*, **206**, 1240, and: Rectification. *Ibid.*, **206**, 1699. Also: Sur les projections d'une loi de probabilité à n variables. *Bull. Sci. Math.*, (2), **63**, 148.

LÉVY, P. (1939*a*). L'addition des variables aléatoires définies sur une circonférence. *Bull. Soc. Math. France*, **67**, 1.

LÉVY, P. (1939*b*). Extensions stochastiques des notions de série, d'intégrale et d'aire. *Comptes rendus*, **209**, 591.

LÉVY, P. (1939*c*). Sur la division d'un segment par les points choisis au hasard. *Comptes rendus*, **208**, 147.

LEWIS, W. T. (1935). A reconsideration of Sheppard's corrections. *Ann. Math. Stats.*, **6**, **11**.

LEXIS, W. (1903). Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik. Fischer, Jena.

Liapounoff, A. (1900). Sur une proposition de la théorie des probabilités. *Bull. Acad. Sci. St. Pét.*, (5), **13**, 359.

Liapounoff, A. (1901). Nouvelle forme du théorème sur la limite de probabilité. *Mem. Acad. Sci. St. Pét.*, (8), **12**, No. 5.

Lidstone, G. T. (1933). Notes on orthogonal polynomials, etc. *J. Inst. Act.*, **64**, 128.

Lidstone, G. T. (1937). Notes on interpolation, etc. *J. Inst. Act.*, **68**, 267.

Lindblad, T. (1937). Zur Theorie der Korrelation bei mehr-dimensionalen zufälligen Variablen. *Acta Soc. Sci. Fennicae*, (2), **10**, 1.

Lindeberg, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Zeit.*, **15**, 211.

Lomnicki, A. (1923). Nouveaux fondements du calcul des probabilités. *Fund. Math.*, **4**, 34.

Lonseth, A. T. (1942). Systems of linear equations with coefficients subject to error. *Ann. Math. Stats.*, **13**, 332.

Lorenz, P. (1931). *Die Trend. Vierteljahrshefte zur Konjunkturforschung.* 2^e Auflage, Sonderhaft, 21.

Lorenz, P. (1935). Annual survey of statistical technique : Trends and seasonal variations. *Econometrika*, **3**, 456.

Lotka, A. J. (1938). Some recent results in population analysis. *J. Am. Stat. Ass.*, **33**, 164.

Lotka, A. J. (1939). On an integral equation in population analysis. *Ann. Math. Stats.*, **10**, 144.

Lüders, R. (1934). Die Statistik der seltenen Ereignisse. *Biom.* **26**, 108.

Lukomski, J. (1939). On some properties of multidimensional distributions. *Ann. Math. Stats.*, **10**, 236.

Lurquin, C. (1937). Sur la loi de Bernouilli à deux variables. *Bull. Classe Sci. Acad. R. Belgique*, (5), **23**, 857.

Macaulay, F. R. (1931). *Smoothing of Time-Series.* National Bureau of Economic Research, New York.

MacMahon, P. A. (1915, 1917). *Combinatory Analysis.* Cambridge University Press.

MacStewart, W. (1941). A note on the power of the sign test. *Ann. Math. Stats.*, **12**, 236.

Madow, W. G. (1937). Contributions to the theory of comparative statistical analysis. I. Fundamental theorems of comparative analysis. *Ann. Math. Stats.*, **8**, 159.

Madow, W. G. (1938). Contributions to the theory of multivariate statistical analysis. *Trans. Am. Math. Soc.*, **44**, 454.

Madow, W. G. (1939). Generalisation of the Laplace-Liapounoff theorem. *Ann. Math. Stats.*, **10**, 84.

Madow, W. G. (1940). Limiting distributions of quadratic and bilinear forms. *Ann. Math. Stats.*, **11**, 125.

Mahalanobis, P. C. (1922). On errors of observation and upper air relationships. *Mem. Ind. Met. Dep.*, **24**.

Mahalanobis, P. C. (1930). On tests and measures of group divergence. I. *J. Asiat. Soc. Beng.*, **26**, 541.

Mahalanobis, P. C. (1933). Tables of L-tests. *Sankhyā*, **1**, 109.

Mahalanobis, P. C., Bose, S. S., Roy, P. R., and Banerji, S. K. (1934). Tables of random samples from a normal population. *Sankhyā*, **1**, 289.

Mahalanobis, P. C. (1936*a*). On the generalised distance in statistics. *Proc. Nat. Inst. Sci. Ind.*, **12**, 49.

Mahalanobis, P. C., Bose, R. C., and Roy, S. N. (1936*b*). Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. *Sankhyā*, **3**, 1.

Mahalanobis, P. C. (1943). An inquiry into the prevalence of drinking tea among middle-class Indian families in Calcutta. *Sankhyā*, **6**, 283.

MÄHLMANN, H. (1935). Ein Beitrag zu Untersuchungen über zweidimensionale Verteilungen von Massenpunkten bei zufallsartig bedingten Bewegungen. *Biom.*, **27**, 191.

MALLOCK, R. R. M. (1933). An electrical calculating machine. *Proc. Roy. Soc.*, **A**, **140**, 457.

MANN, H. B., and WALD, A. (1942). On the choice of the number of intervals in the application of the χ^2-test. *Ann. Math. Stats.*, **13**, 306.

MANN, H. B. (1943). On the construction of sets of orthogonal Latin squares. *Ann. Math. Stats.*, **14**, 401.

MARBE, K. (1934). *Grundfragen der angewandten Wahrscheinlichkeitsrechnung und theoretischer Statistik.* München and Berlin.

MARCH, L. (1926). L'analyse de la variabilité. *Metron*, **6**, No. 2, 3.

MARCHAND, E. (1937). Probabilités expérimentales, probabilités corrigées et probabilités indépendantes. *Mitt. Verein. Schweiz. Versich-Math.*, **33**, 49.

MARCINKIEWICZ, J., and ZYGMUND, A. (1937). Sur les fonctions indépendantes. *Fund. Math.*, **29**, 60.

MARCINKIEWICZ, J. (1939). Sur le problème des moments. *Comptes rendus*, **208**, 405.

MARKOFF, A. A. (1912). *Wahrscheinlichkeitsrechnung.* Teubner, Leipzig.

MARPLES, P. M. (1932). Linear difference equations. *J. Inst. Act.*, **63**, 404.

MARSEGUERRA, V. (1936). Considerazioni sulla cosidetta legge sinusoidale nel calcolo della probabilità. *Giorn. Inst. Ital. Att.*, **7**, 206.

MARTIN, E. S. (1934). On the correction for the moment coefficients of frequency-distributions when the start of the frequency is one of the characteristics to be determined. *Biom.*, **26**, 12.

MARTIN, E. S. (1936). A study of an Egyptian series of mandibles with special reference to mathematical methods of sexing. *Biom.* **28**, 149.

MATHER, K. (1935). The combination of data. *Ann. Eug. Lond.*, **6**, 399.

MATHISEN, H. C. (1943). A method of testing the hypothesis that two samples are from the same population. *Ann. Math. Stats.*, **14**, 188.

MATUSZEWSKI, T., NEYMAN J., and SUPINSKA, J. (1935). Statistical studies in questions of bacteriology. Part I. The accuracy of the dilution method. *Supp. J.R.S.S.*, **2**, 63.

MAZZONI, P. (1934). Su un' origine geometrica di tipi di distribuzioni di frequenze. *Giorn. Ist. Ital. Att.*, **5**, 219.

MCCARTHY, M. D. (1939). On the application of the z-test to randomised blocks. *Ann. Math. Stats.*, **10**, 337.

MCCREA, W. H. (1936). A problem on random paths. *Math. Gaz.*, **20**, 311.

MCKAY, A. T. (1931). Distribution of the estimated coefficient of variation. *J.R.S.S.*, **94**, 564.

MCKAY, A. T. (1932). A Bessel function distribution. *Biom.*, **24**, 39.

MCKAY, A. T., FIELLER, E. C., and PEARSON, E. S. (1932). Distribution of the coefficient of variation and extended t-distribution. *J.R.S.S.*, **95**, 695.

MCKAY, A. T. (1933). The distribution of $\sqrt{\beta_1}$ in samples of four from a normal universe. *Biom.*, **25**, 204; and: The distribution of β_2 in samples of four from a normal universe. *Biom.*, **25**, 411.

MCKAY, A. T., and PEARSON, E. S. (1933). A note on the distribution of range in samples of n. *Biom.*, **25**, 415.

MCKAY, A. T. (1934). Sampling from batches. *Supp. J.R.S.S.*, **1**, 207.

MCKAY, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biom.*, **27**, 466.

MCKINSEY, J. C. C. (1939). A note on Reichenbach's axioms for probability implication. *Bull. Am. Math. Soc.*, **45**, 799.

MCMULLEN, L. (1936). The standard deviation of a difference. *Ann. Eug. Lond.*, **7**, 105.

MEISENER, J. (1938). Erzeugende Funktionen der Charlierschen Polynome. *Math. Zeit.*, **44**, 531.

MENDERSHAUSEN, H. (1937*a*). An example of meaningful curvilinear regression in economic time-series. *Econometrika*, **5**, 329.

MENDERSHAUSEN, H. (1937b). Annual survey of statistical technique: methods of computing and eliminating changing seasonal fluctuations. *Econometrika*, **5**, 234.

MENDERSHAUSEN, H. (1939). Clearing variates in confluence analysis. *J. Am. Stat. Ass.*, **34**, 93.

MENGE, W. O. (1937). A statistical treatment of actuarial functions. *The Record*, **26**, 65.

MERRIL, W. W. (1937). Sampling theory in item analyses. *Psychometrika*, **2**, 215.

MERRINGTON, M. (1941). Numerical approximations to the percentage points of the χ^2-distribution. *Biom.*, **32**, 200.

MERRINGTON, M. (1942). Tables of the percentage points of the t-distribution. *Biom.*, **32**, 300.

MERRINGTON, M., and THOMPSON, C. M. (1943). Tables of the percentage points of the inverted beta (F) distribution. *Biom.*, **33**, 73.

MERZRATH, E. (1933). Anpassung von Flächen an zwei-dimensionale Kollektivgegenstände und ihr Auswerkung für die Korrelationstheorie. *Metron*, **11**, No. 2, 103.

MESSINA, L. (1933). Un teorema sulla legge uniforme dei grandi numeri. *Giorn. Ist. Ital. Att.*, **4**, 116.

MIHOC, G. (1934). Sur les chaînes multiples discontinues. *Comptes rendus*, **198**, 2135.

MIHOC, G. (1935). Sur la détermination de l'intervalle de contraction de la formula de la moyenne. *Comptes rendus*, **200**, 1654.

MILLER, J. C. P. (1934). On a special case in the determination of probable errors. *Month. Not. R. Astr. Soc.*, **94**, 860.

MITCHELL, W. C. (1913). *Business Cycles.* Univ. of California Press, Berkeley.

MITCHELL, W. C., and BURNS, A. F. (1935). The National Bureau's Measures of Cyclical Behaviour. Bull. 57, National Bureau of Economic Research.

MOISSEIEV, N. (1937). Über Stabilitätswahrscheinlichkeitsrechnung. *Math. Zeit.*, **42**, 513.

MOLINA, E. C. (1931). Bayes' theorem. *Ann. Math. Stats.*, **2**, 23.

MOLINA, E. C. (1942). *Tables of Poisson's Exponential Limit.* Van Nostrand Co., Inc., New York.

MOOD, A. M. (1939). On the L_1 test for many samples. *Ann. Math. Stats.*, **10**, 187.

MOOD, A. M. (1940). The distribution theory of runs. *Ann. Math. Stats.*, **11**, 367.

MOOD, A. M. (1943). On the dependence of sampling inspection plans upon population distributions. *Ann. Math. Stats.*, **14**, 415.

MOORE, H. L. (1914). *Economic Cycles: their law and cause.* Macmillan, New York.

MOORE, H. L. (1923). *Generating Economic Cycles.* Macmillan, New York.

MOORE, T. V. (1937). Reduction of data showing non-linear regression for correlation by the ordinary product-moment formula, and the measurement of error due to linear regression. *J. Educ. Psych.*, **28**, 205.

MORANT, G. M. (1921). On random occurrences in space and time when followed by a closed interval. *Biom.* **13**, 309.

MORANT, G. M. (1939). The use of statistical methods in the investigation of problems of classification in anthropology. I. The general nature of the material and the form of intra-racial distributions of metrical characters. *Biom.* **31**, 72.

MORGAN, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biom.*, **31**, 13.

MORTARA, G. (1934). Sulle disuguaglianze statistiche. *22ᵉ Sessione dell' Ist. Int. Stat.* London.

MOSAK, J. L. (1939). The least-square standard error of the coefficient of elasticity of demand. *J. Am. Stat. Ass.*, **34**, 353.

MOULTON, E. J. (1938). The periodic function obtained by repeated accumulation of a statistical series. *Am. Math. Monthly*, **45**, 583.

MOUZON, E. D. (1930). Equimodal frequency distributions. *Ann. Math. Stats.*, **1**, 137.

MUENCH, H. (1936). The probability distribution of protection test results. *J. Am. Stat. Ass.*, **31**, 677.

MUENCH, H. (1938). Discrete frequency-distributions arising from mixtures of several single probability values. *J. Am. Stat. Ass.*, **33**, 390.

Müller, J. H. (1931). On the application of continued fractions to the evaluation of certain integrals, with special reference to the incomplete Beta-function. *Biom.*, **22**, 284.

Musselman, J. R. (1926). On the linear correlation ratio in the case of certain symmetrical frequency-distributions. *Biom.*, **18**, 228.

Myers, R. J. (1934). Note on Koshal's method of improving the parameters of curves by the use of maximum likelihood. *Ann. Math. Stats.*, **5**, 320.

Nagel, E. (1936). The meaning of probability. *J. Am. Stat. Ass.*, **31**, 10.

Nair, A. N. K. (1942). On the distribution of Student's t and the correlation coefficient in samples from non-normal population. *Sankhyā*, **5**, 393.

Nair, K. R. (1936). A note on the extension of lagging correlations between two random series. *J.R.S.S.*, **99**, 559.

Nair, K. R. (1938*a*). On Tippett's random sampling numbers. *Sankhyā*, **4**, 65.

Nair, K. R. (1938*b*). On a method of getting confounded arrangements in the general symmetrical type of experiment. *Sankhyā*, **4**, 121.

Nair, K. R. (1940*a*). The application of covariance technique to field experiments with missing or mixed-up yields. *Sankhyā*, **4**, 581.

Nair, K. R. (1940*b*). Table of confidence intervals for the median in samples from any continuous population. *Sankhyā*, **4**, 551.

Nair, K. R. (1941). Balanced confounded arrangements for the 5^n type of experiment. *Sankhyā*, **5**, 57.

Nair, K. R. (1942). A note on the method of fitting constants for analysis of non-orthogonal data arranged in double classification. *Sankhyā*, **5**, 317.

Nair, K. R., and Rao, C. R. (1942). A note on partially balanced incomplete block designs. *Science and Culture*, **7**, 568.

Nair, K. R., and Shrivastava, M. P. (1942). On a simple method of curve-fitting. *Sankhyā*, **6**, 121.

Nair, K. R. (1943). Certain inequality relationships among the combinatorial parameters of incomplete block designs. *Sankhyā*, **6**, 255.

Nair, K. R., and Banerjee, K. S. (1943). A note on fitting straight lines if both variables are subject to error. *Sankhyā*, **6**, 331.

Nair, U. S. (1936). The standard error of Gini's mean difference. *Biom.*, **28**, 428.

Nair, U. S. (1939). The application of the moment function in the study of distribution laws in statistics. *Biom.*, **30**, 274.

Nair, U. S. (1941*a*). Probability statements regarding the ratio of standard deviations and correlation coefficient in a bivariate normal population. *Sankhyā*, **5**, 151.

Nair, U. S. (1941*b*). A comparison of tests for the significance of the difference between two variances. *Sankhyā*, **5**, 157.

Narumi, S. (1923*a*). On the general forms of bivariate frequency-distributions which are mathematically possible when regression and variation are subjected to limiting conditions. Part I, *Biom.*, **15**, 77, and Part II, *Biom.*, **15**, 209.

Narumi, S. (1923*b*). On further inequalities with possible application to problems in the theory of probability. *Biom.*, **15**, 245.

Nayer, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits. *Stat. Res. Mem.*, **1**, 38.

Newbold, E. M. (1925). Notes on an experimental test of errors in partial correlation coefficients derived from fourfold and biserial total coefficients. *Biom.*, **17**, 251.

Newbold, E. M. (1927). Practical application of the statistics of repeated events, particularly to industrial accidents. *J.R.S.S.*, **90**, 487.

Newland, W. F., and Neal, E. E. (1939). Statistical control of the quality of telephone service. *Supp. J.R.S.S.*, **6**, 25.

NEWMAN, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biom.*, **31**, 20.

NEYMAN, J. (1925). Contributions to the theory of small samples drawn from a finite population. *Biom.*, **17**, 472.

NEYMAN, J. (1926). Further notes on non-linear regression. *Biom.*, **18**, 257.

NEYMAN, J., and PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biom.*, **20A**, 175 and 263.

NEYMAN, J., and PEARSON, E. S. (1931*a*). Further notes on the χ^2-distribution. *Biom.*, **22**, 298.

NEYMAN, J., and PEARSON, E. S. (1931*b*). On the problem of k samples. *Bull. Acad. Polonaise Sci. Lett. Series A*, 460.

NEYMAN, J., and PEARSON, E. S. (1933*a*). On the testing of statistical hypotheses in relation to probability *a priori*. *Proc. Camb. Phil. Soc.*, **29**, 492.

NEYMAN, J. (1933*b*). An outline of the theory and practice of representative method applied in social research. *Polish Inst. Social Problems. Actuarial Series, No.* 1. Warsaw.

NEYMAN, J., and PEARSON, E. S. (1933*c*). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.*, A, **231**, 289.

NEYMAN, J. (1934). On two different aspects of the representative method, etc. *J.R.S.S.*, **97**, 558.

NEYMAN, J. (1935*a*). Su un teorema concernente le cosidette statistiche sufficienti. *Giorn. Ist. Ital. Att.*, **6**, 320.

NEYMAN, J. (1935*b*). Sur la vérification des hypothèses statistiques composées. *Bull. Soc. Math. France*, **63**, 1.

NEYMAN, J., IWASKIEWICZ, K., and KOLODZIECZYK, St., (1935*c*). Statistical problems in agricultural experimentation. *Supp. J.R.S.S.*, **2**, 107.

NEYMAN, J., and PEARSON, E. S. (1936*a*). Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Stat. Res. Mem.*, **1**, 113.

NEYMAN, J., and TOKARSKA, B. (1936*b*). Errors of the second kind in testing ' Student's ' hypothesis. *J. Am. Stat. Ass*, **31**, 318.

NEYMAN, J., and PEARSON, E. S. (1936, 1938). Contributions to the theory of testing statistical hypotheses : I. Unbiassed critical regions of Type A and Type A_1. *Stat. Res. Mem.*, **1**, 1 ; II. Certain theorems on unbiassed critical regions of Type A ; III. Unbiassed tests of simple statistical hypotheses specifying the values of more than one unknown parameter. *Ibid.*, **2**, 25.

NEYMAN, J. (1937*a*). ' Smooth test ' for goodness of fit. *Skand. Akt.*, **20**, 149.

NEYMAN, J. (1937*b*). Outline of a theory of statistical estimation based on the classical theory of probability. *Phil. Trans.*, A, **236**, 333.

NEYMAN, J. (1938*a*). Contribution to the theory of sampling human populations. *J. Am. Stat. Ass.*, **33**, 101.

NEYMAN, J. (1938*b*). Tests of statistical hypotheses which are unbiassed in the limit. *Ann. Math. Stats.*, **9**, 69.

NEYMAN, J. (1938*c*). On statistics the distribution of which is independent of the parameters involved in the original probability law of the observed variables. *Stat. Res. Mem.*, **2**, 58.

NEYMAN, J., and PEARSON, E. S. (1938*d*). Note on some points in ' Student's ' paper on ' Comparison between balanced and random arrangements of field plots.' *Biom.*, **29**, 380.

NEYMAN, J. (1939*a*). On a new class of ' contagious ' distributions applicable in entomology and bacteriology. *Ann. Math. Stats.*, **10**, 35.

NEYMAN, J. (1939*b*). On the hypotheses underlying the applications of statistical methods to routine laboratory analysis. *Ann. Math. Stats.*, **10**, 87.

NEYMAN, J. (1941*a*). Fiducial argument and the theory of confidence intervals. *Biom.*, **32**, 128.

NEYMAN, J. (1941*b*). On a statistical problem arising in routine analysis and in sampling inspections of mass production. *Ann. Math. Stats.*, **12**, 46.

NEYMAN, J. (1942). Basic ideas and some recent results of the theory of testing statistical hypotheses. *J.R.S.S.*, **105**, 292.

NICHOLSON, C. (1941). A geometrical analysis of the frequency-distribution of the ratio between two variables. *Biom.*, **32**, 16.

NICHOLSON, C. (1943). The probability integral for two variables. *Biom.*, **33**, 59.

NORRIS, N. (1935). Inequalities among averages. *Ann. Math. Stats.*, **6**, 27.

NORRIS, N. (1937). Convexity properties of generalised mean value functions. *Ann. Math. Stats.*, **8**, 118.

NORRIS, N. (1938). Some efficient measures of relative dispersion. *Ann. Math. Stats.*, **9**, 214.

NORRIS, N. (1939). The standard errors of the geometric and harmonic means. *Ann. Math. Stats.*, **10**, 84.

NORRIS, N. (1940). The standard errors of the geometric means. *Ann. Math. Stats.*, **11**, 445.

NORTON, H. W. (1937). Use of series in an exact test of significance in a discontinuous distribution. *Ann. Eug. Lond.*, **7**, 349.

NORTON, H. W. (1939). The 7×7 squares. *Ann. Eug. Lond.*, **9**, 269.

NORTON, K. A. (1938). Limits to the accuracy of estimated moment coefficients. *Sankhyā*, **3**, 265.

NYDELL, S. (1919). The mean errors of the characteristics in logarithmic normal distributions. *Skand. Akt.*, **2**, 134.

OLDS, E. G. (1935). Distribution of greatest variates, least variates and intervals of variation in samples from a rectangular universe. *Bull. Am. Math. Soc.*, **41**, 297.

OLDS, E. G. (1937). On the remainder in the approximate evaluation of the probability in the symmetrical case of James Bernoulli's theorem. *Bull. Am. Math. Soc.*, **43**, 806.

OLDS, E. G. (1938*a*). A moment-generating function which is useful in solving certain matching problems. *Bull. Am. Math. Soc.*, **44**, 407.

OLDS, E. G. (1938*b*). Distribution of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Stats.*, **9**, 133.

OLDS, E. G. (1939). Remarks on two methods of sampling inspection. *Ann. Math. Stats.*, **10**, 87.

OLDS, E. G. (1940). On a method of sampling. *Ann. Math. Stats.*, **11**, 355.

OLSHEN, C. A. (1938). Transformations of the Pearson Type III distribution. *Ann. Math. Stats.*, **9**, 176.

ONICESCU, O., and MIHOC, G. L. (1935–1939). Sur les chaînes de variables statistiques. *Comptes rendus*, **200**, 511 ; **202**, 2031 ; L'allure asymptotique de la somme des variables d'une chaîne de Markoff discontinue. *Ibid.*, **205**, 481 ; Sur les sommes de variables enchaînés. *Bull. Math. Soc. Roum. Sci.*, **41**, 99.

OPPENHEIM, S. (1909). Über die Bestimmung der Periode einer periodischer Erscheinung nebst Anwendung auf der Theorie des Erdmagnetismus. *Wien. Sitzber.*, 2*a*, **118**.

O'TOOLE, A. L. (1931, 1932). On symmetric functions and symmetric functions of symmetric functions. *Ann. Math. Stats.*, **2**, 101 and : (Multivariate case), *ibid.*, **3**, 56.

O'TOOLE, A. L. (1933). On the system of curves for which the method of moments is the best method of fitting. *Ann. Math. Stats.*, **4**, 1.

O'TOOLE, A. L. (1934). On the best values of r in samples of R from a finite population of N. *Ann. Math. Stats.*, **5**, 146.

OTTESTAD, P. (1937). On some discontinuous frequency functions and frequency distributions. *Skand. Akt.*, **20**, 75.

OTTESTAD, P. (1939). On the use of the factorial moments in the study of discontinuous frequency distributions. *Skand. Akt.*, **22**, 22.

PAE-TSI-YUAN, (1933). On the logarithmic frequency-distribution and the semi-logarithmic correlation surface. *Ann. Math. Stats.*, **4**, 30.

Pairman, E., and Pearson, K. (1919). On the corrections for moment coefficients of limited-range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range. *Biom.*, **12**, 231.

Palm, C. (1937). Inhomogeneous telephone traffic in full-availability groups. *Ericsson Technics*, No. 1, Stockholm.

Panse, V. G. (1939). Preliminary studies on sampling in field experiments. *Sankhyā*, **4**, 139.

Paulson, E. (1941). On certain likelihood ratio tests associated with the exponential distribution. *Ann. Math. Stats.*, **12**, 301.

Paulson, E. (1942). An approximate normalisation of the analysis of variance distribution. *Ann. Math. Stats.*, **13**, 233.

Pearl, R., and Reed, L. J. (1923). On the mathematical theory of population growth. *Metron*, **3**, No. 1, 6.

Pearl, R. (1930). *Introduction to Medical Biometry and Statistics.* Saunders and Co., Philadelphia and London.

Pearl, R., and Miner, J. R. (1935). On the comparison of groups in respect of a number of measured characters. *Human Biology*, **7**, 95.

Pearl, R. (1937). On the moment product-sums of frequency-distributions. *Human Biology*, **9**, 410.

Pearse, G. E. (1928). On corrections for the moment coefficients of frequency-distributions when there are infinite ordinates at one or both terminals of the range. *Biom.*, **20A**, 314.

Pearson, E. S. (1923). The probable error of a class-index correlation. *Biom.*, **14**, 261.

Pearson, E. S. (1924). Note on the approximations to the probable error of a coefficient of correlation. *Biom.*, **16**, 196.

Pearson, E. S. (1925). Bayes' theorem, examined in the light of experimental sampling. *Biom.*, **17**, 388.

Pearson, E. S. (1926). A further note on the distribution of range in samples taken from a normal population. *Biom.*, **18**, 173.

Pearson, E. S. (1927). Further note on the linear correlation ratio. *Biom.*, **19**, 223.

Pearson, E. S., and Adyanthaya, N. K. (1928, 1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biom.*, **20A**, 356, and **21**, 259.

Pearson, E. S. (1929). Some notes on sampling tests with two variables. *Biom.*, **21**, 337.

Pearson, E. S. (1930). A further development of tests for normality. *Biom.*, **22**, 239.

Pearson, E. S., and Neyman, J. (1930). On the problem of two samples. *Bull. Acad. Polonaise Sci. Lett. Series A*, 73.

Pearson, E. S. (1931*a*, 1932). The test of significance for the correlation coefficient. *J. Am. Stat. Ass.*, **26**, 128, and **27**, 424.

Pearson, E. S. (1931*b*). The analysis of variance in cases of non-normal variation. *Biom.*, **23**, 114.

Pearson, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biom.*, **24**, 404.

Pearson, E. S. (1933*a*). Statistical method in the control and standardisation of the quality of manufactured products. *J.R.S.S.*, **96**, 21.

Pearson, E. S., and Wilks, S. S. (1933*b*). Methods of statistical analysis appropriate for k samples of two variables. *Biom.*, **25**, 353.

Pearson, E. S. (1934). Sampling problems in industry. *Supp. J.R.S.S.*, **1**, 107.

Pearson, E. S., and Haines, J. (1935*a*). The use of range in place of standard deviation in small samples. *Supp. J.R.S.S.*, **2**, 83.

Pearson, E. S., and Sukhatme, A. V. (1935*b*). An illustration of the use of fiducial limits in determining the characteristics of a sampled batch. *Sankhyā*, **2**, 13.

Pearson, E. S. (1935*c*). A comparison of β_2 and Mr. Geary's w_n criterion. *Biom.*, **27**, 333.

PEARSON, E. S. and CHANDRA SEKAR, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biom.*, **28**, 308.

PEARSON, E. S. (1937*a*). Maximum likelihood and methods of estimation. *Biom.*, **29**, 155.

PEARSON, E. S. (1937*b*, 1938). Some aspects of the problem of randomisation. *Biom.*, **29**, **53**. II. An illustration of 'Student's' inquiry into the effect of balancing in agricultural experiments. *Biom.*, **30**, 159.

PEARSON, E. S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biom.*, **30**, 134.

PEARSON, E. S. (1939). Note on the inverse and direct methods of estimation in R. D. Gordon's problem. *Biom.*, **31**, 181.

PEARSON, E. S. (1941). A note on further properties of statistical tests. *Biom.*, **32**, 59.

PEARSON, E. S. (1942*a*). Notes on testing statistical hypotheses. *Biom.*, **32**, 311.

PEARSON, E. S., and HARTLEY, H. O. (1942*b*). The probability integral of the range in samples of *n* observations from a normal population. *Biom.*, **32**, 301.

PEARSON, E. S., and HARTLEY, H. O. (1943). Tables of the probability integral of the 'studentised' range. *Biom.*, **33**, 89.

PEARSON, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans.*, **A**, **185**, 71.

PEARSON, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Phil. Trans.*, **A**, **186**, 343.

PEARSON, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Phil. Trans.*, **A**, **187**, 253.

PEARSON, K., and LEE, A. (1897*a*). On the distribution of frequency (variation and correlation) of the barometric heights at diverse stations. *Phil. Trans.*, **A**, **190**, 423.

PEARSON, K. (1897*b*). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. Roy. Soc.*, **60**, 489.

PEARSON, K., and FILON, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Phil. Trans.*, **A**, **191**, 229.

PEARSON, K., LEE, A., and BRAMLEY-MOORE, L. (1899*a*). Mathematical contributions to the theory of evolution. VI. Genetic (reproduction) selection. Inheritance of fertility in man and of fecundity in thoroughbred racehorses. *Phil. Trans.*, **A**, **192**, 257.

PEARSON, K. (1899*b*). On certain properties of the hypergeometrical series and on the fitting of such series to observation polygons in the theory of chance. *Phil. Mag.*, (5), **47**, 236.

PEARSON, K. (1900*a*). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Phil. Trans.*, **A**, **195**, 1.

PEARSON, K. (1900*b*). Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. *Phil. Trans.*, **A**, **195**, 79.

PEARSON, K. (1900*c*). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. *Phil. Mag.*, (5), **50**, 157.

PEARSON, K., and others (1901*a*). Mathematical contributions to the theory of evolution. IX. On the principle of homotyposis, etc. *Phil. Trans.*, **A**, **197**, 285.

PEARSON, K. (1901*b*). Mathematical contributions to the theory of evolution. X. Supplement to a memoir on skew variation. *Phil. Trans.*, **A**, **197**, 443.

PEARSON, K. (1901*c*). On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, (6), **2**, 559.

PEARSON, K. (1902*a*). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil. Trans.*, **A**, **200**, 1.

PEARSON, K. (1902*b*). On the modal value of an organ or character. *Biom.*, **1**, 256.

PEARSON, K. (1902*c*). Note on Francis Galton's problem. *Biom.*, **1**, 390.

PEARSON, K. (1903, 1913, 1920). On the probable errors of frequency constants. *Biom.*, **2**, 273; **9**, 1 and **13**, 113.

PEARSON, K. (1904). Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation. *Drapers' Co. Res. Mem. Biometric Series I.* Cambridge University Press (formerly Dulau and Co.).

PEARSON, K. (1905). Mathematical contributions to the theory of evolution. XIV. On the general theory of skew correlation and non-linear regression. *Drapers' Co. Res. Mem. Biometric Series II.* Cambridge University Press.

PEARSON, K., and BLAKEMAN, J. (1906). On the probable error of mean-square contingency. *Biom.*, **5**, 191; (PEARSON alone, 1915). *Biom.*, **10**, 570; (with A. W. YOUNG, 1916) On the probable error of a coefficient of contingency without approximation. *Biom.*, **11**, 215.

PEARSON, K. (1907*a*). Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation. *Drapers' Co. Res. Mem. Biometric Series IV.* Cambridge University Press.

PEARSON, K. (1907*b*). On the influence of past experience on future expectation. *Phil. Mag.*, (6), **13**, 365.

PEARSON, K. and LEE, A. (1908). On the generalised probable error in multiple normal correlation. *Biom.*, **6**, 59.

PEARSON, K. (1909). On a new method of determining correlation between a measured character *A* and a character *B*, etc. *Biom.*, **6**, 96.

PEARSON, K. (1910). On a new method of determining correlation when one variable is given by alternative and the other by multiple categories. *Biom.*, **7**, 248.

PEARSON, K. (1911*a*). On the probability that two independent distributions of frequency are really samples from the same population. *Biom.*, **8**, 250.

PEARSON, K. (1911*b*). On a correction to be made to the correlation ratio η. *Biom.*, **8**, 254.

PEARSON, K. (1912*a*). Mathematical contributions to the theory of evolution. XVIII. On a novel method of regarding the association of two variates classed solely in alternate categories. *Drapers' Co. Res. Mem. Biometric Series VII.* Cambridge University Press.

PEARSON, K. (1912*b*, 1913). On the appearance of multiple cases of disease in the same house. *Biom.*, **8**, 404, and **9**, 28.

PEARSON, K. (1913*a*). On the probable error of a coefficient of correlation as found from a fourfold table. *Biom.*, **9**, 22.

PEARSON, K. (1913*b*). On the measurement of the influence of ' broad categories ' on correlation. *Biom.*, **9**, 116.

PEARSON, K. and HERON, D. (1913*c*). On theories of association. *Biom.*, **9**, 159.

PEARSON, K. (1913*d*). Note on the surface of constant association. *Biom.*, **9**, 534.

PEARSON, K., editor, (1914, 1931). *Tables for Statisticians and Biometricians, Part I* (1914, 3rd edn. 1930) and *Part II* (1931). Cambridge University Press.

PEARSON, K., and CAVE, B. M. (1914). Numerical illustrations of the variate-difference correlation method. *Biom.*, **10**, 340.

PEARSON, K. (1914, 1921). On an extension of the method of correlation by grades or ranks. *Biom.*, **10**, 416; and Second note, *Biom.* **13**, 302.

PEARSON, K. (1915*a*). On the partial correlation ratio. *Proc. Roy. Soc.*, **A**, **91**, 492.

PEARSON, K. (1915*b*). On certain types of compound frequency-distributions in which the components can be individually described by binomial series. *Biom.*, **11**, 139.

PEARSON, K. (1916*a*). Mathematical contributions to the theory of evolution. XIX. Second supplement to a memoir on skew variation. *Phil. Trans.*, **A**, **216**, 429. (Correction, *Biom.*, **12**, 259.)

PEARSON, K. (1916*b*). On the general theory of multiple contingency with special reference to partial contingency. *Biom.*, **11**, 145.

PEARSON, K., and TOCHER, J. (1916*c*). On criteria for the existence of differential death-rates. *Biom.*, **11**, 145.

PEARSON, K. (1916*d*). On some novel properties of partial and multiple correlation coefficients in a universe of manifold characteristics. *Biom.*, **11**, 231.

PEARSON, K. (1916*e*). On the application of 'goodness of fit' tables to test regression curves (and theoretical curves) used to describe observational or experimental data. *Biom.*, **11**, 239. (Correction, *Biom.*, **12**, 259.)

PEARSON, K. (1916*f*). On a brief proof of the fundamental formula for testing the goodness of fit of frequency-distributions and on the probable error of *P*. *Phil. Mag.*, (6), **31**, 369.

PEARSON, K. (1917). On the probable error of biserial η. *Biom.*, **11**, 292.

PEARSON, K. and YOUNG, A. W. (1918). On the product-moments of various orders of the normal correlation surface of two variates. *Biom.*, **12**, 86.

PEARSON, K., and PAIRMAN, E. (1919). On corrections for the moment-coefficients of limited range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range. *Biom.*, **12**, 231.

PEARSON, K. (1919). On generalised Tchebycheff theorems in the mathematical theory of statistics. *Biom.*, **12**, 284.

PEARSON, K. (1920*a*). The fundamental problem of practical statistics. *Biom.*, **13**, 1.

PEARSON, K. (1920*b*). Notes on the history of correlation. *Biom.*, **13**, 25.

PEARSON, K. (1920*c*). *On the Construction of Tables and on Interpolation. Part I. Univariate Tables. Part II. Bivariate Tables.* Tracts for Computers, Nos. 2 and 3. Cambridge University Press.

PEARSON, K. (1921). On a general method of determining the successive terms in a skew regression line. *Biom.*, **13**, 296.

PEARSON, K. (1922*a*, 1923). On the χ^2 test of goodness of fit. *Biom.*, **14**, 186; and: Further note, *Biom.*, **14**, 418.

PEARSON, K., and PEARSON, E. S. (1922*b*). On polychoric coefficients of correlation. *Biom.*, **14**, 127.

PEARSON, K., and ELDERTON, E. M. (1923*a*). On the variate-difference method. *Biom.*, **14**, 281.

PEARSON, K. (1923*b*). On the correction necessary for the correlation ratio η. *Biom.*, **14**, 412.

PEARSON, K. (1923*c*). Notes on skew frequency surfaces. *Biom.*, **15**, 222; and: On non-skew frequency surfaces. *Biom.*, **15**, 231.

PEARSON, K. (1924*a*). Note on Professor Romanovsky's generalisation of my frequency curves. *Biom.*, **16**, 116.

PEARSON, K. (1924*b*). On the moments of the hypergeometrical series. *Biom.*, **16**, 157.

PEARSON, K. (1924*c*) On a certain double hypergeometrical series and its representation by continuous frequency surfaces. *Biom.*, **16**, 172.

PEARSON, K. (1924*d*). On the mean error of frequency-distributions. *Biom.*, **16**, 198.

PEARSON, K. (1924*e*). Historical note on the origin of the normal curve. *Biom.*, **16**, 402.

PEARSON, K. (1925*a*). The fifteen-constant bivariate frequency surface. *Biom.*, **17**, 268.

PEARSON, K. (1925*b*). On first-power methods of finding correlation. *Biom.*, **17**, 459.

PEARSON, K. (1926*a*). Researches on the mode of distribution of the constants of samples taken at random from a bivariate normal population. *Proc. Roy. Soc.*, **A**, **112**, 1.

PEARSON, K. (1926*b*). On the coefficient of racial likeness. *Biom.*, **18**, 105.

PEARSON, K., JEFFERY, G. B., and ELDERTON, E. M. (1929). On the distribution of the first product-moment coefficient in samples drawn from an indefinitely large normal population. *Biom.*, **21**, 164.

PEARSON, K. (1931*a*). On the nature of the relationship between two of 'Student's' variates (z_1 and z_2) when samples are taken from a bivariate normal population. *Biom.*, **22**, 405;

Some properties of 'Student's' z. *Biom.*, **23**, 1; and: Further remarks on the z-test. *Biom.*, **23**, 408.

PEARSON, K. (1931*b*). Appendix to a paper by Professor Tokishige Hojo. On the standard error of the median to a third approximation, etc. *Biom.*, **23**, 361.

PEARSON, K., and PEARSON, M. V. (1931*c*, 1932). On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. Part I. Symmetrical Distributions (Normal and Rectangular). *Biom.*, **23**, 364. Part II. Case of certain skew curves. *Biom.*, **24**, 203.

PEARSON, K. (1931*d*). Historical note on the distribution of the standard deviation of samples of any size drawn from an indefinitely large normal parent population. *Biom.*, **23**, 416.

PEARSON, K., STOUFFER, S. A., and DAVID, F. N. (1932*a*). Further applications in statistics of the $T_m(x)$ Bessel function. *Biom.*, **24**, 293.

PEARSON, K. (1932*b*). Experimental discussion of the (χ^2, P) test for goodness of fit. *Biom.*, **24**, 351.

PEARSON, K. (1933*a*). On the applications of the double Bessel function $K_{\tau_1\tau_2}(x)$ to statistical problems. Part I. Theoretical. *Biom.*, **25**, 158.

PEARSON, K. (1933*b*). On a method of determining whether a sample of given size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biom.*, **25**, 379.

PEARSON, K. (1934). On a new method of determining 'goodness of fit'. *Biom.*, **26**, 425.

PEARSON, K. (1935). On the corrections for broad categories, being a note on Mr. Wisniewski's memoir. *Biom.*, **27**, 364.

PEARSON, K. (1936). Method of moments and method of maximum likelihood. *Biom.*, **28**, 34.

PEEK, R. L. (1937). Test of an observed difference in the frequency of two results. *J. Am. Stat. Ass.*, **32**, 532.

PEIERLS, R. S. (1935). Statistical error in counting experiments. *Proc. Roy. Soc.*, **A**, **149**, 467.

PEISER, M. A. (1943). Asymptotic formulæ for significance levels of certain distributions. *Ann. Math. Stats.*, **14**, 56.

PEPPER, J. (1929). Studies in the theory of sampling. *Biom.*, **21**, 231.

PEPPER, J. (1932). The sampling distribution of the third moment coefficient—an experiment. *Biom.*, **24**, 55.

PERLO, V. (1933). On the distribution of 'Student's' ratio for samples of three drawn from a rectangular population. *Biom.* **25**, 203.

PERSONS, W. M. (1928). *The construction of index numbers*. Houghton Mifflin, Cambridge, Mass.

PIETRA, G. (1925). The theory of statistical relations, with special reference to cyclical series. *Metron*, **4**, No. 3–4, 383.

PIETRA, G. (1932*a*). Nuovi contributi alla metodologia degli indici di variabilità e di concentrazione. *Att. R. Ist. Veneto di Sci.*, 989.

PIETRA, G. (1932*b*). Dell' interpolazione parabolica nel caso in cui entrambi i valori delle variabili sono affetti da errori accidentali. *Metron*, **9**, Nos. 3–4, 77.

PIETRA, G. (1934). *Statistica.* (2 vols.) Giuffrè, Milan.

PITMAN, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.*, **32**, 567.

PITMAN, E. J. G. (1937*a*, 1938). Significance tests which may be applied to samples from any population. *Supp. J.R.S.S.*, **4**, 119; II. The correlation coefficient test. *Supp. J.R.S.S.* **4**, 225; III. The analysis of variance test. *Biom.*, **29**, 322.

PITMAN, E. J. G. (1937*b*). The 'closest' estimates of statistical parameters. *Proc. Camb. Phil. Soc.*, **33**, 212.

PITMAN, E. J. G. (1939*a*). The estimation of location and scale parameters of a continuous population of any given form. *Biom.*, **30**, 391.

PITMAN, E. J. G. (1939*b*). Tests of hypotheses concerning location and scale parameters. *Biom.*, **31**, 200.

PITMAN, E. J. G. (1939*c*). A note on normal correlation. *Biom.*, **31**, 9.

PIZZETTI, E. (1939). Osservazioni sulle medie esponenziali e baso-esponenziali. *Metron*, **13**, No. 4, 3.

POISSON, S. D. (1837). *Recherches sur la probabilité des jugements, etc.* Paris.

POLLAK, L. W. (1926). *Rechentafeln zur harmonischen Analysis.* Barth, Leipzig.

POLLAK, L. W. (1927). Periodogramme hochfrequenten Schwankungen meteorologischer Elemente. *Met. Zeit.*, **4**, 121.

POLLAK, L. W., and KAISER, F. (1935). Méthode numérique de J. Fuhrich pour le calcul des périodicités, sa mise à l'épreuve et son application aux mouvements polaires. *Rév. Stat. Tchéchoslovaque*, **16**, 13.

POLLARD, H. S. (1934). On the relative stability of the median and arithmetic mean, with particular reference to certain frequency-distributions which can be dissected into normal distributions. *Ann. Math. Stats.*, **5**, 227.

PÓLYA, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentproblem. *Math. Zeit.*, **8**, 173.

PÓLYA, G. (1923). Herleitung des Gauss'schen Gesetzes aus einer Funktionalgleichung. *Math. Zeit.*, **18**, 96.

PÓLYA, G. (1931). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré*, **1**, 117.

PÓLYA, G. (1937). Zur Kinematik der Geschiebebewegung. *Mitt. Versuchs. Wasserbau an der Eid. Tech. Hochschule.* Zurich.

PÓLYA, G. (1938*a*). Sur l'indétermination d'un problème voisin du problème des moments. *Comptes rendus*, **207**, 708.

PÓLYA, G. (1938*b*). Sur la promenade au hasard dans un réseau de rues. *Actualités Scientifiques et Industrielles*, No. 734, Paris. Hermann et Cie.

POWELL, R. W. (1930). Successive integration as a method for finding long-period cycles. *Ann. Math. Stats.*, **1**, 123.

PRETORIUS, S. J. (1930). Skew bivariate frequency surfaces examined in the light of numerical illustrations. *Biom.*, **22**, 109.

PROKOPOVIC, S. N. (1935). La corrélation des séries quantitatives. *Rev. Stat. Tchéchoslovaque*, **16**, 64.

PRZYBOROWSKI, J. and WILÉNSKI, H. (1935*a*). Sur les erreurs de la première et de la seconde catégorie dans la vérification des hypothèses concernant la loi de Poisson. *Comptes rendus*, **200**, 1460.

PRZYBOROWSKI, J., and WILÉNSKI, H. (1935*b*). Statistical principles of routine work in testing clover seed for dodder. *Biom.*, **27**, 273.

PRZYBOROWSKI, J., and WILÉNSKI, H. (1936). Note on the application of a theorem of Frau Pollaczek-Geiringer. *Biom.*, **28**, 187.

PRZYBOROWSKI, J., and WILÉNSKI, H. (1940). Homogeneity of results in testing samples from Poisson series, etc. *Biom.*, **31**, 313.

QUENSEL, C. E. (1936). A method of determining the regression curve when the marginal distribution is of the normal logarithmic type. *Ann. Math. Stats.*, **7**, 196.

QUENSEL, C. E. (1938). The distributions of the second moment and of the correlation coefficient in samples from populations of Type A. *Lunds Univ. Arsskr.*, N.F. **34**, 4, 1.

RAIKOV, D. (1938). On the decomposition of Gauss and Poisson laws. *Bull. Acad. Sci. U.S.S.R. Sér. Math.*, **1**, 91.

REED, L. J. (1922). Fitting straight lines. *Metron*, **1**, No. 3, 54.

REGAN, F. (1936, 1938). The application of the theory of admissible numbers to time-series with constant probability. *Trans. Am. Math. Soc.*, **36**, 511 ; and : The application of the theory of admissible numbers to time-series with variable probability. *Am. J. Maths.*, **58**, 867.

Reichenbach, H. (1935). *Wahrscheinlichkeitslehre,* Leiden.

Reichenbach, H. (1937). Les fondements logiques du calcul des probabilités. *Ann. Inst. H. Poincaré,* **7**, 267.

Reiersöl, O. (1940). A method for recurrent computation of all the principal minors of a determinant and its application in confluence analysis. *Ann. Math. Stats.,* **11**, 193.

Reiersöl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrika,* **9**, 1.

Rhodes, E. C. (1921). *Smoothing.* Tracts for Computers, No. 6. Cambridge University Press.

Rhodes, E. C. (1923, 1925). On a certain skew correlation surface. *Biom.,* **14**, 355, and **17**, 314.

Rhodes, E. C. (1924, 1925). On the problem whether two given samples can be supposed to have been drawn from the same population. *Biom.,* **16**, 239 and *Metron,* **5**, 3.

Rhodes, E. C. (1925). On sampling. *Metron,* **5**, Nos. 2–3, 3.

Rhodes, E. C. (1927). The precision of means and standard deviations when the individual errors are correlated. *J.R.S.S.,* **90**, 135.

Rhodes, E. C. (1928). On the normal correlation function as an approximation to the distribution of paired drawings. *J.R.S.S.,* **91**, 548.

Rhodes, E. C. (1930). On the fitting of parabolic curves to statistical data. *J.R.S.S.,* **93**, 569.

Rhodes, E. C. (1936). The precision of index numbers. *J.R.S.S.,* **99**, 142.

Rice, S. O. (1938). Van Uven's theorem in probability theory and a self-reciprocal Hankel transform. *Quart. J. Maths,* **9**, 1.

Rice, S. O. (1939). The distribution of the maxima of a random curve. *Am. J. Maths.,* **61**, 409.

Richards, H. I. (1931). Analysis of the spurious effect of high intercorrelation of independent variables on regression and correlation coefficients. *J. Am. Stat. Ass.,* **26**, 21.

Ricker, W. E. (1937). The concept of confidence or fiducial limits applied to the Poisson frequency. *J. Am. Stat. Ass.,* **32**, 349.

Rider, P. R. (1929). On the distribution of the ratio of mean to standard deviation in small samples from non-normal populations. *Biom.,* **21**, 124.

Rider, P. R. (1931*a*). On small samples from certain non-normal universes. *Ann. Math. Stats.,* **2**, 48.

Rider, P. R. (1931*b*). A note on small sample theory. *J. Am. Stat. Ass.,* **26**, 172.

Rider, P. R. (1932). On the distribution of the correlation coefficient in small samples. *Biom.,* **24**, 382.

Rider, P. R. (1933). Criteria for rejection of observations. *Washington University Studies, New Series, Science and Technology, No.* 8.

Rider, P. R. (1934). The third and fourth moments of the generalised Lexis theory. *Metron,* **12**, No. 1, 185.

Rider, P. R. (1936). Annual survey of statistical technique : Developments in the analysis of multivariate data. *Econometrika,* **4**, 264.

Rietz, H. L., editor (1924). *Handbook of Mathematical Statistics.* Houghton Mifflin, Boston.

Rietz, H. L. (1931*a*). Note on the distribution of the standard deviation of sets of three variates drawn at random from a rectangular distribution. *Biom.,* **23**, 424.

Rietz, H. L. (1931*b*). On certain properties of frequency-distributions obtained by a linear fractional transformation of the variates of a given distribution. *Ann. Math. Stats.,* **2**, 38.

Rietz, H. L. (1932). A simple non-normal correlation surface. *Biom.,* **24**, 288.

Rietz, H. L. (1937). Some topics in sampling theory. *Bull. Am. Math. Soc.,* **43**, 209.

Rietz, H. L. (1938). On a recent advance in statistical inference. *Am. Math. Monthly,* **45**, 149.

Rietz, H. L. (1939). On the distribution of the 'Student' ratio for small samples from certain non-normal populations. *Ann. Math. Stats.,* **10**, 265.

Risser, R. (1935–7). Exposé des principes de la statistique mathématique. *J. Soc. Stat. Paris,* **76**, 281 ; **77**, 337 ; **78**, 40.

Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biom.*, **12**, **93**.

Ritchie-Scott, A. (1921). The incomplete moments of a normal solid. *Biom.*, **13**, 401.

Robb, R. A. (1929). The variate-difference method of seasonal variation. *J. Am. Stat. Ass.*, **24**, 250.

Robb, R. A. (1930). Modifications of the link relative and interpolation methods of determining seasonal variation. *Ann. Math. Stats.*, **1**, 352.

Robinson, S. (1933). An experiment regarding the χ^2-test. *Ann. Math. Stats.*, **4**, 285.

Roff, M. (1937). Relation between results obtainable with raw and corrected correlation coefficients in multiple factor analysis. *Psychometrika*, **2**, 35.

Romanovsky, V. (1923). Note on the moments of a binomial about its mean. *Biom.*, **15**, **410**.

Romanovsky, V. (1924). Generalisation of some types of the frequency-curves of Professor Pearson. *Biom.*, **16**, 106.

Romanovsky, V. (1925*a*). On the moments of standard deviation and of correlation coefficient in samples from normal. *Metron*, **5**, No. 4, 3.

Romanovsky, V. (1925*b*). On the moments of the hypergeometrical series. *Biom.*, **17**, 57.

Romanovsky, V. (1926). On the distribution of the regression coefficient in samples from normal population. *Bull. Acad. Sci. U.S.S.R.*, (6), **10**, 643.

Romanovsky, V. (1927). Note on orthogonalising series of functions and interpolation. *Biom.*, **19**, 93.

Romanovsky, V. (1928). On the criteria that two given samples belong to the same normal population. *Metron*, **7**, No. 3, 3.

Romanovsky, V. (1929). On the moments of means of functions of one and more random variables. *Metron*, **8**, Nos. 1–2, 251.

Romanovsky, V. (1931*a*). *Sulla probabilità a posteriori*. *Giorn. Ist. Ital. Att.*, **2**.

Romanovsky, V. (1931*b*). Sulle regressione multiple. *Giorn. Ist. Ital. Att.*, **2**.

Romanovsky, V. (1931*c*, 1932*a*, 1933*a*). Généralisations d'un théorème de M. Slutzky. *Comptes rendus*, **192**, 718 ; Sur la loi sinusoïdale limite. *Rend. Circ. Mat. Palermo*, **56**, 1 ; Sur une généralisation de la loi sinusoïdale limite. *Ibid.*, **57**.

Romanovsky, V. (1932*b*). Due nuovi criteri di controllo sull' andamento casuale di una successione di valori. *Giorn. Ist. Ital. Att.*, **3**, 203.

Romanovsky, V. (1933*b*). On a property of the mean ranges in samples from a normal population and on some integrals of Professor Hojo. *Biom.*, **25**, 195.

Romanovsky, V. (1934). Su due problemi di distribuzione casuale. *Giorn. Ist. Ital. Att.*, **5**, 196.

Romanovsky, V. (1936*a*). Recherches sur les chaînes de Markoff. *Acta Math.*, **66**, 147.

Romanovsky, V. (1936*b*). Note on the method of moments. *Biom.*, **28**, 188.

Romanovsky, V. (1938). Analytical inequalities and statistical tests. (Russian, with English summary). *Bull. Acad. Sci. U.S.S.R. Sér. Math.*, **4**, 457.

Roos, C. F. (1934). *Dynamic Economics*. Bloomington, Indiana.

Roos, C. F. (1936). Annual survey of statistical technique. The correlation and analysis of time-series. *Econometrika*, **4**, 368.

Roos, C. F. (1937). A general invariant criterion of fit for lines and planes where all variates are subject to error. *Metron*, **13**, No. 1, 3.

Roy, S. N. (1938). A geometrical note on the use of rectangular co-ordinates in the theory of sampling distributions connected with a multivariate normal population. *Sankhyā*, **3**, 273.

Roy, S. N. (1939*a*) A note on the distribution of the Studentised D^2-statistic. *Sankhyā*, **4**, 373.

Roy, S. N. (1939*b*). p-statistics, or some generalisations on analysis of variance appropriate to multivariate problems. *Sankhyā*, **4**, 381.

Roy, S. N. (1942*a*). The sampling distribution of p-statistics and certain allied statistics on the non-null hypothesis. *Sankhyā*, **6**, 15.

Roy, S. N. (1942*b*). Analysis of variance for multivariate normal populations, etc. *Sankhyā*, **6**, 35.

Salvemini, T. (1934). Ricerche sperimentali sull' interpolazione grafica di istogrammi. *Metron*, **11**, No. 4, 83.

Salvemini, T. (1939). L' indice di cograduazione del Gini nel caso di serie statistiche con ripetizioni. *Metron*, **13**, No. 4, 27.

Salvosa, L. R. (1930). Tables of Pearson's Type III function. *Ann. Math. Stats.*, **1**, 191.

Samuelson, P. A. (1942). A method of determining explicitly the coefficients of a characteristic equation. *Ann. Math. Stats.*, **13**, 424.

Samuelson, P. A. (1943). Fitting general Gram-Charlier series. *Ann. Math. Stats.*, **14**, 179.

Sandon, F. (1924). Note on the simplification of the calculation of abruptness coefficients to correct crude moments. *Biom.*, **16**, 193.

Sansone, G. (1933). La chiusura dei sistemi ortogonali di Legendre, di Laguerre e di Hermite rispetto alle funzioni di quadrati sommabili. *Giorn. Ist. Ital. Att.*, **4**, 71.

Sasuly, M. (1934). *Trend Analysis of Statistics*. Brookings Institution, Washington, D.C.

Satterthwaite, F. E. (1943). Generalised Poisson distribution. *Ann. Math. Stats.*, **13**, 410.

Savur, S. R. (1937*a*). The use of the median in tests of significance. *Proc. Indian Acad. Sci.*, **A**, **5**, 564.

Savur, S. R. (1937*b*). A new solution of a problem in inverse probability. *Proc. Indian Acad. Sci.*, **A**, **5**, 222.

Savur, S. R. (1939). A note on the arrangement of incomplete blocks when $k = 3$ and $\lambda = 1$. *Ann. Eug. Lond.*, **9**, 45.

Savur, S. R. (1941). A test of significance in approximate periodogram analysis. *Sankhyā*, **6**, 77.

Scheffé, H. (1942*a*). On the theory of testing composite hypotheses with one constraint. *Ann. Math. Stats.*, **13**, 280.

Scheffé, H. (1942*b*). On the ratio of the variances of two normal samples. *Ann. Math. Stats.*, **13**, 371.

Scheffé, H. (1943). Statistical inference in the non-parametric case. *Ann. Math. Stats.*, **14**, 305.

Schmidt, E. (1934). Über die Charlier-Jordansche Entwicklung einer willkürlicher Funktion nach der Poissonsche Funktion und ihrer Ableitungen. *Zeit. ang. Math. und Mech.*, **13**, 139.

Schmidt, R. (1934). Statistical analysis of one-dimensional distributions. *Ann. Math. Stats.*, **5**, 30.

Schultz, H. (1930). The standard error of a forecast from a curve. *J. Am. Stat. Ass.*, **25**, 139.

Schultz, H. (1933). The standard error of the coefficient of elasticity of demand. *J. Am. Stat. Ass.*, **28**, 64.

Schultz, H. (1939). A misunderstanding in index number theory : the true Konös condition on cost-of-living index numbers and limitations. *Econometrika*, **7**, 1.

Schultz, T. W., and Snedecor, E. (1933). Analysis of variance as an effective method of handling the time element in certain economic statistics. *J. Am. Stat. Ass.*, **28**, 14.

Schumann, T. E. W. (1938). A general graduation formula for the smoothing of time series. *Phil. Mag.*, **26**, 970.

Schumann, T. E. W., and Hofmeyer, W. L. (1942). The problem of auto-correlation of meteorological time-series, etc. *Q.J. Met. Soc.*, **68**, 177.

Schuster, Sir Arthur (1898). On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena. *Terr. Mag.*, **3**, 13.

Schuster, Sir Arthur (1899). The periodogram of the magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871–1895. *Trans. Camb. Phil. Soc.*, **18**, 107.

Schuster, Sir Arthur (1906). On the periodicities of sunspots. *Phil. Trans.*, **A**, **206**, 69.

Sčukarev, A. N. (1932). Über die Mechanik der Massenprozesse (Kollektivgegenstandlehre). *Metron*, **9**, Nos. 3–4, 139.

Seal, H. L. (1940). Tests of a mortality table graduation. *J. Inst. Act.*, **71**, No. 330.

SEGAL, J. E. (1938). Fiducial distribution of several parameters with application to a normal system. *Proc. Camb. Phil. Soc.*, **34**, 41.

SHEFFER, H. M. (1935). Concerning some methods of best approximation and a theorem of Birkhoff. *Am. J. Maths.*, **57**, 587.

SHEPPARD, W. F. For bibliography see *Ann. Eug. Lond.*, 1937, **8**, 13.

SHEPPARD, W. F. (1898*a*). On the application of the theory of error to cases of normal distributions and normal correlations. *Phil. Trans.*, **A**, **192**, 101, and *Proc. Roy. Soc.*, **62**, 170.

SHEPPARD, W. F. (1898*b*). On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale. *Proc. Lond. Math. Soc.*, **29**, 353.

SHEPPARD, W. F. (1914). Fitting of polynomials by the method of least squares. *Proc. Lond. Math. Soc.*, (2), **13**, 97.

SHEPPARD, W. F. (1929). The fit of a formula for discrepant observations. *Phil. Trans.*, **A**, **228**, 199.

SHEPPARD, W. F. (1939, posthumous). *The Probability Integral.* British Ass. Math. Tables, vol. 7. Cambridge University Press.

SHEWHART, W. A., and WINTERS, F. W. (1928). Small samples—new experimental results. *J. Am. Stat. Ass.*, **23**, 144.

SHEWHART, W. A. (1931). *The Economic Control of Quality of a Manufactured Product.* van Nostrand, New York.

SHOHAT, J. (1929). Inequalities for moments of frequency functions and for various statistical constants. *Biom.*, **21**, 361.

SHOHAT, J. (1930). Stieltjes integrals in mathematical statistics. *Ann. Math. Stats.*, **1**, **73**.

SHOHAT, J. (1935). On the development of functions in series of orthogonal polynomials. *Bull. Am. Math. Soc.*, **41**, 49.

SIMAIKA, J. B. (1941). On an optimum property of two important statistical tests. *Biom.*, **32**, 70.

SIMAIKA, J. B. (1942). Interpolation for fresh probability levels between the standard table levels of a function. *Biom.*, **32**, 263.

SIMON, H. A. (1943). Symmetric tests of the hypothesis that the mean of one normal population exceeds that of another. *Ann. Math. Stats.*, **14**, 149.

SIMON, L. E. (1941). *The Engineer's Manual of Statistical Methods.* John Wiley, New York.

SIMONSEN, W. (1937). On the distributions of certain functions of samples from a multivariate infinite population. *Skand. Akt.*, **20**, 200.

SIPOS, A. (1930). *Practical application of Jordan's method for trend measurement.* Hornyansky, Budapest.

SLUTZKY, E. (1914). On the criterion of goodness of fit of regression lines and on the best method of fitting them to data. *J.R.S.S.*, **77**, 78.

SLUTZKY, E. (1925). Über stochastische Asymptoten und Grenzwerte. *Metron,* **5**, No. 3, 3.

SLUTZKY, E. (1934). Alcune applicazioni dei coefficienti di Fourier all' analisi delle funzioni aleatorie stazionarie. *Giorn. Ist. Ital. Att.*, **5**, 435.

SLUTZKY, E. (1937*a*). Qualche proposizione relativa alla teoria delle funzioni aleatorie. *Giorn. Ist. Ital. Att.*, **8**, 183.

SLUTZKY, E. (1937*b*). The summation of random causes as the source of cyclic processes. *Econometrika*, **5**, 105.

SMIRNOFF, N. (1935). Über die Verteilung des allgemeinen Gliedes in der Variationsreihe. *Metron,* **12**, No. 2, 59.

SMIRNOFF, N. (1936). Sur la distribution de ω^2. *Comptes rendus*, **202**, 449.

SMITH, C. D. (1930). On generalised Tchebycheff inequalities in mathematical statistics. *Am. J. Maths.*, **52**, 109.

SMITH, C. D. (1939). On Tchebycheff approximations for decreasing functions. *Ann. Math. Stats.* **10**, 190.

Smith, H. Fairfield (1936). A discriminant function for plant selection. *Ann. Eug. Lond.*, **7**, 240.

Smith, K. (1916). On the 'best' values of the constants in frequency-distributions. *Biom.*, **11**, 262.

Smith, K. (1918). On the standard deviation of the adjusted and interpolated values of an observed polynomial function and its constants etc. *Biom.*, **12**, 1.

Smith, K. (1922). The standard deviations of fraternal and parental correlation coefficients. *Biom.*, **14**, 1.

Snedecor, G. W., and Irwin, M. R. (1933). On the chi-square test for homogeneity. *Iowa State College J. Sci.*, **8**, 75.

Snedecor, G. W., and Cox, G. M. (1934*a*). Disproportionate sub-class numbers in tables of multiple classification. *Iowa Agr. Exp. Station Res. Bull.* No. 180.

Snedecor, G. W. (1934*b*). *Calculation and Interpretation of Analysis of Variance and Covariance.* Collegiate Press, Ames, Iowa.

Snedecor, G. W. (1935). Analysis of covariance of statistically controlled grades. *J. Am. Stat. Ass.*, **30**, *Supp.*, 263.

Snow, E. C. (1911). On restricted lines and planes of closest fit to systems of points in any number of dimensions. *Phil. Mag.* (6), **21**, 367.

Solomon, R. S. (1939). An index of conformity based on the *J*-curve hypotheses. *Sociometry*, **2**, 63.

Soper, H. E. (1913). On the probable error of the correlation coefficient to a second approximation. *Biom.*, **9**, 91.

Soper, H. E. (1914). On the probable error of the bi-serial expression for the correlation coefficient. *Biom.*, **10**, 384.

Soper, H. E. and others (1917). On the distribution of the correlation coefficient in small samples. *Biom.*, **11**, 328.

Soper, H. E. (1922). *Frequency Arrays.* Cambridge University Press.

Soper, H. E. (1929*a*). The general sampling distribution of the multiple correlation coefficient. *J.R.S.S.*, **92**, 445.

Soper, H. E. (1929*b*). The interpretation of periodicity in disease prevalence. *J.R.S.S.*, **92**, 34.

'Sophister' (1928). Discussion of small samples from an infinite skew universe. *Biom.*, **20A**, 389.

Spearman, C. (1906). A footrule for measuring correlation. *Brit. J. Psych.*, **2**, 89.

Spearman, C. (1907). Demonstration of formulæ for true measurement of correlation. *Am. J. Psych.*, **18**, 161.

Spearman, C. (1910). Correlation calculated from faulty data. *Brit. J. Psych*, **3**, 271.

Starkey, D. M. (1938). A test of significance of the difference between means of samples from two normal populations without assuming equal variances. *Ann. Math. Stats.*, **9**, 201.

Starkey, D. M. (1939). The distribution of the multiple correlation coefficient in periodogram analysis. *Ann. Math. Stats.*, **10**, 327.

Steffensen, J. F. (1923). *Matematisk Iagtagelseslaehre.* Copenhagen.

Steffensen, J. F. (1930). *Some Recent Researches in the Theory of Statistics and Actuarial Science.* Cambridge University Press.

Steffensen, J. F. (1934). On certain measures of dependence between statistical variables. *Biom.*, **26**, 251.

Steffensen, J. F. (1936). Free functions and the Student–Fisher theorem. *Skand. Akt.*, **19**, 108.

Steffensen, J. F. (1937). On the semi-normal distribution. *Skand. Akt.*, **20**, 60.

Steinhaus, H. (1923). Les probabilités dénombrables et leur rapport à la théorie de la mesure. *Fund. Math.*, **4**, 286.

Stekloff, W. (1914). Quelques applications nouvelles de la théorie de fermeture au problème de représentation approchée de fonctions et au problème des moments. *Mem. Acad. Imp. Sci. St. Pét.*, **32**, No. 4.

STERNE, T. E. (1934). The accuracy of least-square solutions. *Proc. Nat. Acad. Sci.*, **20**, 565 and 601.

STEVENS, W. L. (1936). The analysis of interference. *J. Genetics*, **32**, 51.

STEVENS, W. L. (1937*a*). The truncated normal distribution. *Ann. Appl. Biol.*, **24**, 847.

STEVENS, W. L. (1937*b*). Significance of grouping. *Ann. Eug. Lond.*, **8**, 57.

STEVENS, W. L. (1938*a*). The distribution of entries in a contingency table with fixed marginal totals. *Ann. Eug. Lond.*, **8**, 238.

STEVENS, W. L. (1938*b*). The completely orthogonalised Latin square. *Ann. Eug. Lond.*, **9**, 82.

STEVENS, W. L. (1939*a*). Solution to a geometrical problem in probability. *Ann. Eug. Lond.*, **9**, 315.

STEVENS, W. L. (1939*b*). Tests of significance for extra-sensory perception data. *Psych. Rev.*, **46**, 142.

ST. GEORGESCU, N. (1932). Further contributions to the sampling problem. *Biom.*, **24**, 65.

STIELTJES, J. (1918). Recherches sur les fractions continues. *Œuvres*, Groningen.

STOCK, J. S., and FRANKEL, L. R. (1939). The allocation of samplings among several strata. *Ann. Math. Stats.*, **10**, 288.

STOUFFER, S. A., and TIBBITS, C. (1933). Tests of significance in applying Westergaard's method of expected cases to sociological data. *J. Am. Stat. Ass.*, **28**, 293.

STOUFFER, S. A. (1934). A coefficient of combined partial correlation with an example from sociological data. *J. Am. Stat. Ass.*, **29**, 70.

STOUFFER, S. A. (1936*a*). Evaluating the effect of inadequately measured variables in partial correlation analysis. *J. Am. Stat Ass.*, **31**, 348.

STOUFFER, S. A. (1936*b*). Reliability coefficients in a correlation matrix. *Psychometrika*, **1**, 17.

'STUDENT' (W. S. GOSSET) (1907). On the error of counting with a hæmacytometer. *Biom.*, **5**, 351.

'STUDENT' (1908*a*). On the probable error of a mean. *Biom.*, **6**, 1.

'STUDENT' (1908*b*). On the probable error of a correlation coefficient. *Biom.*, **6**, 302.

'STUDENT' (1909). On the distribution of means of samples which are not drawn at random. *Biom.*, **7**, 210.

'STUDENT' (1913). The correction to be made to the correlation ratio for grouping. *Biom.*, **9**, 316.

'STUDENT' (1914). The elimination of spurious correlation due to position in time or space. *Biom.*, **10**, 179.

'STUDENT' (1919). An explanation of deviations from Poisson's law in practice. *Biom.*, **12**, 211.

'STUDENT' (1921). An experimental determination of the probable error of Dr. Spearman's correlation coefficient. *Biom.*, **13**, 263.

'STUDENT' (1927). Errors of routine analysis. *Biom.*, **19**, 151.

'STUDENT' (1931*a*). On the z-test. *Biom.*, **23**, 407.

'STUDENT' (1931*b*). Yield trials. Article in Baillière's *Encyclopedia of Scientific Agriculture*.

'STUDENT' (1931*c*). The Lanarkshire milk experiment. *Biom.*, **23**, 398.

'STUDENT' (1938). Comparison between balanced and random arrangements of field plots. *Biom.*, **29**, 361.

STUMPFF, K. (1926). Fehlertheoretische Untersuchungen zur Periodogrammanalyse. *Astr. Nach.*, **226**, 378.

STUMPFF, K. (1937). *Grundlagen und Methoden der Periodenforschung*. Berlin.

SUBRAMANIAN, S. (1935). On a property of partial correlation. *J.R.S.S.*, **98**, 129.

SUKHATME, P. V. (1935). Contribution to the theory of the representative method. *Supp. J.R.S.S.*, **2**, 253.

SUKHATME, P. V. (1936*a*). A contribution to the problem of two samples. *Proc. Indian Acad. Sci.*, **2**, **A**, 584.

SUKHATME, P. V. (1936*b*). On the analysis of k samples from exponential populations with especial reference to the problem of random intervals. *Stat. Res. Mem.*, **1**, 94.

SUKHATME, P. V. (1937*a*). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Ann. Eug. Lond.*, **8**, 52.

SUKHATME, P. V. (1937*b*). The problem of k samples for Poisson population. *Proc. Nat. Inst. Sci. India*, **3**, 297.

SUKHATME, P. V. (1938*a*). On the distribution of χ^2 in samples from a Poisson series. *Supp. J.R.S.S.*, **5**, 75.

SUKHATME, P. V. (1938*b*). On Fisher and Behrens' test of significance for the difference in means of two normal samples. *Sankhyā*, **4**, 39.

SUKHATME, P. V. (1938*c*). On bipartitional functions. *Phil. Trans. Roy. Soc.*, **A**, **237**, 375.

SUKHATME, P. V. (1944). Moments and product-moments of moment-statistics for samples of the finite and infinite populations. *Sankhyā*, **6**, 363.

SWAROOP, S. (1938). Tables of the exact values of probabilities for testing the significance of differences between proportions based on pairs of small samples. *Sankhyā*, **4**, 73.

SWED, F. S., and EISENHART, C. (1943). Tables for testing randomness in a sequence of alternatives. *Ann. Math. Stats.*, **14**, 66.

TANG, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Stat. Res. Mem.*, **2**, 126.

TANG, Y. (1938). Certain statistical problems arising in plant breeding. *Biom.*, **30**, 29.

TAPPAN, M. (1927). On partial multiple correlation coefficients in a universe of manifold characteristics. *Biom.*, **19**, 39.

TARTLER, A. (1935). On a certain class of orthogonal polynomials. *Am. J. Maths.*, **57**, 627.

TAUCER, R. (1936). I fenomeni di selezione a la teoria dei gruppi. *Giorn. Ist. Ital. Att.*, **7**, 16.

TCHEBYCHEFF, P. L. (1907) *Œuvres.* 2 vols. St. Pétersbourg: including: Sur une formule d'analyse, **1**, 701 (1854); Sur les fractions continues, **1**, 203 (1856); Sur une nouvelle série, **1**, 381 (1858); Sur l'interpolation par la méthode des moindres carrés, **1**, 478 (1859); Sur l'interpolation des valeurs équidistantes, **2**, 219 (1875).

TEDIN, O. (1931). The influence of systematic plot arrangements upon the estimate of error in field experiments. *J. Agr. Sci.*, **21**, 191.

THIELE, T. N. (1931). *Theory of Observations.* Reprint in *Ann. Math. Stats.*, **2**, 165, of the English version published in 1903.

THOMPSON, C. M., PEARSON, E. S., COMRIE, L. J., and HARTLEY, H. O. (1941*a*). Tables of percentage points of the incomplete beta-function. *Biom.*, **32**, 151.

THOMPSON, C. M. (1941*b*). Tables of percentage points of the χ^2-distribution. *Biom.*, **32**, 187.

THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biom.*, **25**, 286.

THOMPSON, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sampling standard deviation. *Ann. Math. Stats.*, **6**, 214.

THOMPSON, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Ann. Math. Stats.*, **7**, 122.

THOMPSON, W. R. (1938). Biological applications of normal range and associated significance tests in ignorance of original distribution forms. *Ann. Math. Stats.*, **9**, 281.

THOMSON, G. H. (1916). A hierarchy without a general factor. *Brit. J. Psych*, **8**, 271.

THOMSON, G. H. (1919*a*). The criterion of goodness of fit of psychophysical curves. *Biom.*, **12**, 216.

THOMSON, G. H. (1919*b*). On the degree of perfection of hierarchical order among correlation coefficients. *Biom.*, **12**, 355, and (correction), **15**, 150.

THOMSON, G. H. (1935). On complete families of correlation coefficients, etc. *Brit. J. Psych.*, **26**, 63.

THOMSON, G. H. (1939). The factorial analysis of ability. *Brit. J. Psych.*, **30**, 71 and 105.

THORNDIKE, E. L. (1937). On correlations between measurements which are not normally distributed. *J. Educ. Psych.*, **28**, 367.

THOULESS, R. H. (1939). The effect of errors of measurement on correlation coefficients. *Brit. J. Psych*, **29**, 383.

THURSTONE, L. L. (1935). *Vectors of Mind.* Chicago.

THURSTONE, L. L. (1938). A new rotational method in factor analysis. *Psychometrika*, **43**, **199**.

TINBERGEN, J. (1937). *An Econometric Approach to Business Cycle Problems.* Paris.

TINBERGEN, J. (1938). On the theory of business cycle control. *Econometrika*, **6**, 22.

TINTNER, G. (1935). *Prices and the Trade Cycle.* Vienna.

TINTNER, G. (1940). *The Variate-Difference Method.* Bloomington Press, Indiana.

TINTNER, G. (1941). The variate-difference method: a reply. *Econometrika*, **9**, 163.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biom.*, **17**, 364.

TIPPETT, L. H. C. (1931). *The Methods of Statistics.* Williams and Norgate, London. 2nd edn. 1937.

TIPPETT, L. H. C. (1932). A modified method of counting particles. *Proc. Roy. Soc.*, **A**, **137**, 434.

TIPPETT, L. H. C. (1935). Some applications of statistical methods to the study of variation of quality in the production of cotton yarn. *Supp. J.R.S.S.*, **2**, 27.

TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace.* Macmillan, London.

TORNIER, E. (1929). Wahrscheinlichkeitsrechnung und Zahlentheorie. *J. rein. und ang. Math.*, **160**, 177.

TORNIER, E. (1930). Die Axiome der Wahrscheinlichkeitsrechnung. *J. rein. und ang. Math.*, **163**, 45.

TORNIER, E. (1933). Grundlagen der Wahrscheinlichkeitsrechnung. *Acta Math.*, **60**, 239.

TORNIER, E. (1936). *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie.* Teubner, Leipzig.

TORNIER, E. (1937). Verallgemeinerung des Rückschluss-Satzes der Wahrscheinlichkeitsrechnung. *Deutsche Math.*, **2**, 469.

TRACHTENBERG, H. L. (1921). Analysis of the periodogram. *J.R.S.S.*, **84**, 578.

TRAVERS, R. M. W. (1939). The use of a discriminant function in the treatment of psychological group-differences. *Psychometrika*, **4**, 25.

TRELOAR, A. E., and WILDER, M. A. (1934). The adequacy of 'Student's' criterion of deviations in small sample means. *Ann. Math. Stats.*, **5**, 324.

TRICOMI, F. (1935, 1936*a*). Su la rappresentazione di una legge di probabilità mediante esponenziali di Gauss e la transformazione di Laplace. *Giorn. Ist. Ital. Att.*, **6**, 135 and **7**, 42.

TRICOMI, F. (1936*b*). Sulla media dei valori assoluti di errori seguente la legge di Gauss. *Giorn. Ist. Ital. Att.*, **7**, 280.

TRICOMI, F. (1937). Sul rapporto fra la media dei quadrati di più errori e il quadrato della media dei loro valori assoluti. *Giorn. Ist. Ital. Att.*, **8**, 68 and 127.

TRICOMI, F. (1938). Les transformations de Fourier, Laplace, Gauss et leurs applications au calcul des probabilités et à la statistique. *Ann. Inst. H. Poincaré*, **8**, 111.

TRUKSA, L. (1940). The simultaneous distribution in samples of mean and standard deviation and of mean and variance. *Biom.*, **31**, 256.

TSCHUPROW, A. A. (1918, 1919*a*). Zur Theorie der Stabilität statistischer Reihen. *Skand. Akt.*, **1**, 199 (1918), and **2**, 80 (1919).

TSCHUPROW, A. A. (1918*b*, 1921, 1923). On the mathematical expectation of the moments of frequency-distributions. *Biom.*, **12**, 140 and 185; *Biom.*, **13**, 283; and *Metron*, **2**, No. 3, 461 and No. 4, 646.

TSCHUPROW, A. A. (1925). *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Teubner, Leipzig. (English translation as *The Mathematical Theory of Correlation*, William Hodge, 1939.)

TSCHUPROW, A. A., trans. by L. ISSERLIS (1928). The mathematical theory of the statistical

methods employed in the study of correlation in the case of three variables. *Trans. Camb. Phil. Soc.*, **23**, 337.

TSCHUPROW, A. A. (1934). The mathematical foundations of the methods to be used in statistical investigation of the dependence between two chance variables. *Nordisk Statistik Tidskrift*, **5**, 34.

TURNER, H. H. (1913). *Tables for facilitating the use of harmonic analysis.* Oxford University Press.

URBAN, F. M. (1918). Über den Begriff der mathematischen Wahrscheinlichkeit. *Vierteljahrschrift für Wiss. Phil. and Soz.*, **10**.

USPENSKY, J. V. (1937). *Introduction to Mathematical Probability*, McGraw-Hill, New York and London.

VAJDA, S. (1939). Die Wahrscheinlichkeit einer bestimmten Auszahlungssumme. *Skand. Akt.*, **22**, 10.

VAN DER POL, B. (1930). Oscillations sinusoïdales et de rélaxation. (*L'onde électrique*, juin-juillet, Chiron, Paris.

VAN KAMPEN, E. R. (1937*a*). On the addition of convex curves and the densities of certain infinite convolutions. *Am. J. Maths.* **59**, 679.

VAN KAMPEN, E. R. and WINTNER, A. (1937*b*, 1937*c*). Convolutions of distributions on convex curves and the Riemann zeta-function. *Am. J. Maths.*, **59**, 175; and: On divergent infinite convolutions. *Ibid.* **59**, 635.

VAN KAMPEN, E. R. (1939*a*). On the asymptotic distribution of a uniformly almost periodic function. *Am. J. Maths.* **61**, 729.

VAN KAMPEN, E. R. and WINTNER, A. (1939*b*). A limit theorem for probability distributions on lattices. *Am. J. Maths.* **61**, 965.

VAN UVEN, M. J. (1932). Compensazione degli errori di un rapporto. *Metron*, **10**, No. 3, 185.

VAN UVEN, M. J. (1939). Adjustment of a ratio. *Ann. Eug. Lond.*, **9**, 181.

VENN, J. A. (1888). *The Logic of Chance.* 3rd edn., Macmillan, London. (Out of print.)

VERNON, P. E. (1936). A note on the standard error in the contingency-matching technique. *J. Educ. Psych.*, **27**, 704.

VILLARS, D. S., and ANDERSON, T. W. (1943). Some significance tests for normal bivariate distributions. *Ann. Math. Stats.*, **14**, 141.

VILLE, J. A. (1936*a*, *b*). Sur les suites indifférentes. *Comptes rendus*, **202**, 1393; and: Sur la notion de collectif. *Ibid.*, **203**, 26.

VILLE, J. A. (1936*c*). Sur la convergence de la médiane des n premiers résultats d'une suite infinie d'épreuves indépendantes. *Comptes rendus*, **203**, 1309.

VILLE, J. A. (1939). Étude critique de la notion de collectif. Paris: Thèse.

VINCI, F. (1920). Sui coefficienti di variabilità. *Metron*, **1**, No. 1, 62.

VINCI, F. (1934). Significant developments in business cycle theory. *Econometrika*, **2**, 125.

VOLTERRA, V. (1936). Les équations des fluctuations biologiques et le calcul des variations. *Comptes rendus*, **202**, 1935; Les équations canoniques. *Ibid.*, **202**, 2023; and: Sur l'intégration des équations. *Ibid.*, **202**, 2113.

VON BORTKIEWICZ, L., (1898). *Das Gesetz der kleinen Zahlen.* Teubner, Leipzig.

VON BORTKIEWICZ, L. (1910). Zur Verteidigung des Gesetzes der kleinen Zahlen. *Jahrb. Nat. Ök. und Stat.*, (3), **39**, 218.

VON BORTKIEWICZ, L. (1915*a*). Über die Zeitfolge zufälliger Ereignisse. *Bull. Inst. Int. de Stat.*, **20**, 2e livre.

VON BORTKIEWICZ, L. (1915*b*). Realismus und Formalismus in der mathematischen Statistik. *Allg. Stat. Arkiv.*, **9**, 225.

VON BORTKIEWICZ, L. (1917). *Die Iterationen.* Berlin.

VON BORTKIEWICZ, L. (1922). Das Helmertsche Verteilunggesetz für die Quadratsumme zufälliger Beobachtungsfehler. *Zeit. ang. Math. und Mech.* **2**, Heft 5.

VON BORTKIEWICZ, L. (1931). The relation between stability and homogeneity. *Ann. Math. Stats.* **2**, 1.

VON MISES, R. (1919*a*, *b*). Fundamentalsätze der Wahrscheinlichkeitsrechnung. *Math. Zeit.*, **4**, 1 and: Grundlagen, *ibid.*, **5**, 52.

VON MISES, R. (1921). Das Problem der Iterationen. *Zeit. ang. Math. und Mech.*, **1**, 298.

VON MISES, R. (1928). *Wahrscheinlichkeit, Statistik und Wahrheit.* Springer, Berlin; 3rd rev. edn., 1936; trans. as *Probability, Statistics and Truth,* 1939. W. Hodge, London.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung.* Deuticke, Wien.

VON MISES, R. (1933). Über Zahlenfolge die ein Kollektivähnliches Verhalten zeigen. *Math. Ann.*, **108**, 757.

VON MISES, R. (1936*a*). Sul concetto di probabilità fondato sul limite di frequenze relative. *Giorn. Ist. Ital. Att.*, **7**, 235.

VON MISES, R. (1936*b*). Les lois de probabilité pour les fonctions statistiques. *Ann. Inst. H. Poincaré*, **6**, 185.

VON MISES, R. (1937). Bestimmung einer Verteilung durch ihre ersten Momente. *Skand. Akt.*, **20**, 220.

VON MISES, R. (1938). A modification of Bayes' problem. *Ann. Math. Stats.*, **9**, 256.

VON MISES, R. (1939*a*). The limits of a distribution function if two expected values are given. *Ann. Math. Stats.*, **10**, 99.

VON MISES, R. (1939*b*). An inequality for the moments of a discontinuous distribution. *Skand. Akt.*, **22**, 32.

VON MISES, R. (1939*c*). Über Aufteilungs- und Besitzungs-Wahrscheinlichkeiten. *Rev. Fac. Sci. Univ. Istamboul*, (4), Fasc. 1–2, 145.

VON MISES, R. (1941). On the foundations of probability and statistics. *Ann. Math. Stats*, **12**, 191.

VON NEUMANN J., and others (1941*a*, *b*). The mean-square successive difference. *Ann. Math. Stats.*, **12**, 153; (VON NEUMANN alone): Distribution of the ratio of the mean-square successive difference to the variance. *Ibid.*, **12**, 367; VON NEUMANN and HART, B. I.: Tabulation of the probabilities for the ratio of the mean-square successive difference to the variance. *Ibid.*, **13**, 207.

VON SCHELLING, H. (1934). Die Konzentration einer Verteilung und ihre Abhängigkeit von den Grenzen des Variationsbereiches. *Metron*, **11**, No. 4, 3.

VON SZELISKI, V. S. (1929). Experiments in the correlation of time-series. *J. Am. Stat. Ass.*, **24**, *Supp.*, 241.

WALD, A. (1936*a*). *Berechnung und Ausschaltung von Saisonschwankungen.* Beiträge zur Konjunkturforschung. *Oester. Inst. Konj.*, **9**, 35.

WALD, A. (1936*b*). Sur la notion de collectif dans le calcul des probabilités. *Comptes rendus*, **202**, 180.

WALD, A. (1937). Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung. *Ergeb. math. Kolloqu., Hamburg*, No. 8, 38.

WALD, A. (1938). A generalisation of Markoff's inequality. *Ann. Math. Stats.*, **9**, 244.

WALD, A. (1939*a*). Contributions to the theory of statistical estimation and testing hypotheses. *Ann. Math. Stats.*, **10**, 299.

WALD, A., and WOLFOWITZ, J. (1939*b*). Confidence limits of continuous distribution functions. *Ann. Math. Stats.*, **10**, 105.

WALD, A. (1940*a*). The fitting of straight lines if both variables are subject to error. *Ann. Math. Stats.*, **11**, 284.

WALD, A. (1940*b*). A note on the analysis of variance with unequal class-frequencies. *Ann. Math. Stats.*, **11**, 96.

WALD, A. and WOLFOWITZ, J. (1940*c*). On a test whether two samples are from the same population. *Ann. Math. Stats.*, **11**, 147.

WALD, A. (1941*a*). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Stats.*, **12**, 1 and 396.

WALD, A., and BROOKNER, R. J. (1941*b*). On the distribution of Wilks' statistic, etc. *Ann. Math. Stats.*, **12**, 137.

WALD, A., and WOLFOWITZ, J. (1941*c*). Note on confidence limits for continuous distribution functions. *Ann. Math. Stats.*, **12**, 118.

WALD, A. (1941*d*). On the analysis of variance in case of multiple classifications with unequal class frequencies. *Ann. Math. Stats.*, **12**, 346.

WALD, A. (1942*a*). Asymptotically shortest confidence intervals. *Ann. Math. Stats.*, **13**, 127.

WALD, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Stats.*, **14**, 134.

WALKER, SIR GILBERT (1914). On the criterion for the reality of relationships or periodicities. *Calcutta Ind. Met. Mems.*, **21**, part 9.

WALKER, SIR GILBERT (1925). On periodicity. *Q. J. Roy. Met. Soc.*, **51**, 387.

WALKER, SIR GILBERT (1927). On periodicity and its existence in European weather. *Mem. Roy. Met. Soc.*, **1**, No. 9.

WALKER, SIR GILBERT (1931). On periodicity in series of related terms. *Proc. Roy. Soc.*, **A**, **131**, 518.

WALKER, H. M. (1929). *Studies in the History of Statistical Method.* Williams and Wilkins, Baltimore.

WALKER, H. M., and SANFORD, V. (1934). The accuracy of computation with approximate numbers. *Ann. Math. Stats.*, **5**, 1.

WALLACE, N., and TRAVERS, R. M. W. (1938). A psychometric sociological study of a group of speciality salesmen. *Ann. Eug. Lond.*, **8**, 266.

WALLIS, W. A. (1939). The correlation ratio for ranked data. *J. Am. Stat. Ass.*, **34**, 533.

WALLIS, W. A., and MOORE, G. H. (1941). *A significance test for time-series.* Technical Paper No. 1. Nat. Bur. Ec. Research.

WALLIS, W. A. (1942). Compounding probabilities from independent significance tests. *Econometrika*, **10**, 229.

WATKINS, G. P. (1933). An ordinal index of correlation. *J. Am. Stat. Ass.*, **28**, 139.

WAUGH, F. V. (1942). Regressions between sets of variables. *Econometrika*, **10**, 290.

WEBSTER, M. S. (1938). Orthogonal polynomials with orthogonal derivations. *Bull. Am. Math. Soc.*, **44**, 880.

WEIDA, F. M. (1934). On measures of contingency. *Ann. Math. Stats.*, **5**, 308.

WEIDA, F. M. (1935). On certain distribution functions when the law of the universe is Poisson's first law of error. *Ann. Math. Stats.*, **6**, 102.

WEISS, M. G., and COX, G. M. (1939). Balanced incomplete block and lattice-square designs for testing yield differences among large numbers of soya bean varieties. *Iowa Agr. Exp. Stat. Res. Bull.*, **257**, 289.

WELCH, B. L. (1935). Some problems in the analysis of regression among k samples of two variables. *Biom.*, **27**, 145.

WELCH, B. L. (1936*a*). Note on an extension of the L_1 test. *Stat. Res. Mem.*, **1**, 52.

WELCH, B. L. (1936*b*). Specification of rules for rejecting too variable a product, etc. *Supp. J.R.S.S.*, **3**, 29.

WELCH, B. L. (1937). On the z-test in randomised blocks and Latin squares. *Biom.*, **29**, 21.

WELCH, B. L. (1938*a*). On tests for homogeneity. *Biom.*, **30**, 149.

WELCH, B. L. (1938*b*). The significance of the difference between two means when the population variances are unequal. *Biom.*, **29**, 350.

WELCH, B. L. (1939*a*). On confidence limits and sufficiency with particular reference to parameters of location. *Ann. Math. Stats.*, **10**, 58.

WELCH, B. L. (1939*b*). Note on discriminant functions. *Biom.*, **31**, 218.

WELCH, B. L. (1939*c*). On the distribution of maximum likelihood estimates. *Biom.*, **31**, 187.

WERTHEIMER, A. (1932). A generalised error function. *Ann. Math. Stats.*, **3**, 64.

WERTHEIMER, A. (1937). Note on Zoch's paper on the postulate of the arithmetic mean. *Ann. Math. Stats.*, **8**, 112.

WHERRY, R. J. (1935). The shrinkage of the Brown–Spearman prophecy formula. *Ann. Math. Stats.*, **6**, 183.

WHITAKER, L. (1914). On Poisson's law of small numbers. *Biom.*, **10**, 36.

WHITTAKER, E. T., and ROBINSON, G. (1940). *The Calculus of Observations.* 3rd edn. Blackie & Sons.

WHITWORTH, W. A. (1901). Choice and Chance. 5th edn. Deighton Bell and Co. Cambridge.

WICKSELL, S. D. (1917*a*). On logarithmic correlation with an application to the distribution of ages at first marriage. *Medd. Lunds Astr. Obs.*, No. 84.

WICKSELL, S. D. (1917*b*). The correlation function of Type A. *Kungl. Svenska Vetenskapsakad. Handl.* Bd. 58; *Medd. Lunds Astr. Obs.* Series 2, No. 17.

WICKSELL, S. D. (1921). An exact formula for spurious correlation. *Metron,* **1**, No. 4, 33.

WICKSELL, S. D. (1933). On correlation functions of Type III. *Biom.*, **25**, 121.

WICKSELL, S. D. (1934*a*). Expansions of frequency functions for integer variates in series. *Skand. Matematikercongressen i Stockholm,* p. 306.

WICKSELL, S. D. (1934*b*). Analytical theory of regression. *Medd. Lunds Astr. Obs.* Series 2, No. 69.

WIDDER, D. V. (1934). The inversion of the Laplace integral and the related moment problem. *Trans. Am. Math. Soc.*, **36**, 107.

WIENER, N. (1930). Generalised harmonic analysis. *Acta Math.*, **55**, 117.

WIENER, N. (1938). The homogeneous chaos. *Am. J. Math.*, **60**, 897.

WILKS, S. S. (1932*a*). Moments and distributions of estimates of population parameters from fragmentary samples. *Ann. Math. Stats.*, **3**, 163.

WILKS, S. S. (1932*b*). On the sampling distribution of the multiple correlation coefficient. *Ann. Math. Stats.*, **3**, 196.

WILKS, S. S. (1932*c*). On the distribution of statistics in samples from a normal population of two variables with matched sampling for one variable. *Metron,* **9**, Nos. 3–4, 87.

WILKS, S. S. (1932*d*). The standard error of a tetrad in samples from a normal population of independent variables. *Proc. Nat. Acad. Sci.*, **18**, 562.

WILKS, S. S. (1932*e*). Certain generalisations in the analysis of variance. *Biom.*, **24**, 471.

WILKS, S. S. (1934). Moment-generating operators for determinants of product-moments in samples from a normal system. *Ann. Math.*, **35**, 312.

WILKS, S. S. (1935*a*). The likelihood test of independence in contingency tables. *Ann. Math. Stats.*, **6**, 190.

WILKS, S. S. (1935*b*) On the independence of k sets of normally distributed statistical variables. *Econometrika*, **3**, 309.

WILKS, S. S. (1935*c*). Test criteria for statistical hypotheses involving several variables. *J. Am. Stat. Ass.*, **30**, 549.

WILKS, S. S. (1936). The sampling theory of systems of variances, covariances and intra-class covariances. *Ann. J. Math.*, **58**, 426.

WILKS, S. S., and THOMPSON, C. M. (1937*a*). The sampling distribution of the criterion λ_{H_1} when the hypothesis tested is not true. *Biom.*, **29**, 124.

WILKS, S. S. (1937*b*). The analysis of variance for two or more variables. *Third Ann. Conf. Econ. Stat.* Colorado Springs, p. 82.

WILKS, S. S. (1938*a*). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stats.*, **9**, 60.

WILKS, S. S. (1938*b*). Shortest average confidence intervals from large samples. *Ann. Math. Stats.*, **9**, 166.

WILKS, S. S. (1938*c*). Fiducial distributions in fiducial inference. *Ann. Math. Stats.*, **9**, 272.

WILKS, S. S. (1938*d*). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, **3**, 23.

WILKS, S. S. (1938*e*). The analysis of variance and covariance in non-orthogonal data. *Metron*, **13**, No. 2, 141.

WILKS, S. S. (1939*a*). Optimum fiducial regions for simultaneous estimation of several population parameters for large samples. (Abstract). *Ann. Math. Stats.*, **10**, 85.

WILKS, S. S., and DALY, J. F. (1939*b*). An optimum property of confidence regions associated with the likelihood function. *Ann. Math. Stats.*, **10**, 225.

WILKS, S. S. (1941). On the determination of sample sizes for setting tolerance limits. *Ann. Math. Stats.*, **12**, 91.

WILKS, S. S. (1943). *Mathematical Statistics.* Princeton University Press.

WILLIAMS, C. B. (1937). The use of logarithms in the interpretation of certain entomological problems. *Ann. App. Biol.*, **24**, 404.

WILLIAMS, J. D. (1941). Moments of the ratio of the mean-square successive difference to the mean-square difference in samples from a normal population. *Ann. Math. Stats.*, **12**, 239.

WILSDON, B. H. (1934). Discrimination by specification statistically considered and illustrated by the standard specification for Portland cement. *Supp. J.R.S.S.*, **1**, 152.

WILSON, E. B. (1928). On hierarchical correlation systems. *Proc. Nat. Acad. Sci.*, **14**, 283.

WILSON, E. B., and HILFERTY, M. M. (1931*a*). The distribution of chi-square. *Proc. Nat. Acad. Sci.*, **17**, 694.

WILSON, E. B., HILFERTY, M. M., and MAHER, H. C. (1931*b*). Goodness of fit. *J. Am. Stat. Ass.*, **26**, 443.

WILSON, E. B. (1938). The standard deviation of sampling for life expectancy. *J. Am. Stat. Ass.*, **33**, 705.

WINTNER, A. (1934*a*). On the addition of independent distributions. *Am. J. Maths.*, **56**, 8.

WINTNER, A. (1934*b*). On the asymptotic differential distribution of almost periodic and related functions. *Am. J. Maths.*, **56**, 401.

WINTNER, A. (1935). Papers on convergent convolutions. *Am. J. Maths.*, **57**, 363, 821, 827, 839; and *Bull. Am. Math. Soc.*, **41**, 137.

WINTNER, A. (1936). On a class of Fourier transforms. *Am. J. Maths.*, **58**, 45.

WISHART, J. (1926). On Romanovsky's generalised frequency curves. *Biom.*, **18**, 221.

WISHART, J. (1927). On the approximate quadrature of certain skew curves with an account of the researches of Thomas Bayes. *Biom.*, **19**, 1.

WISHART, J. (1928). The generalised product-moment distribution in samples from a normal multivariate population. *Biom.*, **20A**, 32.

WISHART, J. (1929*a*). The correlation between product-moments of any order in samples from a normal population. *Proc. Roy. Soc. Edin.*, **49**, 1.

WISHART, J. (1929*b*). A problem of combinatorial analysis giving the distribution of certain moment-statistics. *Proc. Lond. Math. Soc.*, **29**, 309.

WISHART, J. (1930). The derivation of certain high-order sampling product-moments from a normal population. *Biom.*, **22**, 224.

WISHART, J. (1931*a*). Notes on frequency constants. *J. Inst. Act.*, **62**, 174.

WISHART, J. (1931*b*). The mean and second-moment coefficient of the multiple correlation coefficient in samples from a normal population. *Biom.*, **22**, 353.

WISHART, J. (1932*a*). A note on the distribution of the correlation ratio. *Biom.*, **24**, 441.

WISHART, J., and BARTLETT, M. S. (1932*b*). The distribution of second-order moment coefficients in small samples. *Proc. Camb. Phil. Soc.*, **28**, 455.

WISHART, J. (1933*a*). The theory of orthogonal polynomial fitting. *J.R.S.S.*, **96**, 487.

WISHART, J. (1933*b*). A comparison of the semi-invariants of the distributions of moment and semi-invariant estimates in samples from an infinite population. *Biom.*, **25**, 52.

WISHART, J., and BARTLETT, M. S. (1933*c*). The generalised product-moment distribution in a normal system. *Proc. Camb. Phil. Soc.*, **29**, 260.

WISHART, J. (1934*a*). Statistics in agricultural research. *Supp. J.R.S.S.*, **1**, 26.

WISHART, J. (1934*b*). Bibliography of agricultural statistics. *Supp. J.R.S.S.*, **1**, 95.

WISHART, J., and SANDERS, H. G. (1935). *Principles and Practice of Field Experimentation.* Empire Cotton-growing Corporation, London.

WISHART, J. (1936). Tests of significance in analysis of covariance. *Supp. J.R.S.S.*, **3**, 79.

WISHART, J. (1938). Field experiments of factorial design. *J. Agr. Sci.*, **28**, 299.

WISHART, J. (1939). Statistical treatment of animal experiments. *Supp. J.R.S.S.*, **6**, 1.

WISNIEWSKI, J. (1934). Interdependence of cyclical and seasonal fluctuation. *Econometrika*, **2**, 176.

WISNIEWSKI, J. (1935, 1936). On the validity of a certain Pearson's formula. *Biom.*, **27**, 356; and: Rejoinder. *Biom.*, **28**, 190.

WISNIEWSKI, J. (1937*a*). A problem in least squares. *Ann. Math. Stats.*, **8**, 145.

WISNIEWSKI, J. (1937*b*). A note on inverse probability. *J.R.S.S.*, **100**, 417.

WOLD, H. (1934*a*). Sulle correzione di Sheppard. *Giorn. Ist. Ital. Att.*, **4**, 304.

WOLD, H. (1934*b*). Sheppard's correction formulæ in several variables. *Skand. Akt.*, **17**, 248.

WOLD, H. (1935). A study on the mean difference, concentration curves and concentration ratio. *Metron*, **12**, No. 2, 39.

WOLD, H. (1936). On quantitative statistical analysis. *Skand. Akt.*, **19**, 281.

WOLD, H. (1938*a*). *A Study in the Analysis of Stationary Time-Series.* Almquist and Wiksells, Uppsala.

WOLD, H. (1938*b*). On the inversion of moving averages. *Skand. Akt.*, **21**, 208.

WOLD, H. (1939). Über stochastische Prozesse, inbesondere solche stationärer Natur. 9 *Cong. des Math. Scand. Helsingfors*, p. 207.

WOLFOWITZ, J. (1942). Additive partition functions and a class of statistical hypotheses. *Ann. Math. Stats.*, **13**, 247.

WOLFOWITZ, J. (1943). On the theory of runs with some applications to quality control. *Ann. Math. Stats.*, **14**, 280.

WONG, Y. K. (1935). An application of the orthogonalisation process to the theory of least squares. *Ann. Math. Stats.*, **6**, 53.

WONG, Y. K. (1937). On the elimination of variables in multiple correlation. *J. Am. Stat. Ass.*, **32**, 357.

WOODBURY, M. A. (1940). Rank correlation when there are equal variates. *Ann. Math. Stats.*, **11**, 358.

WORKING, H. and HOTELLING, H. (1929). Applications of the theory of error to the interpretation of trends. *J. Am. Stat. Ass.*, **24**, *Supp.*, 73.

WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Stats.*, **5**, 161.

YASUKAWA, K. (1925). On the means, standard deviations, correlations and frequency-distributions of functions of variates. *Biom.*, **17**, 211.

YASUKAWA, K. (1926). On the probable error of the mode of frequency-distributions. *Biom.*, **18**, 263.

YASUKAWA, K. (1934). On the deviation from normality of the frequency-distributions of functions of normally distributed variates. *Tokohu Math. J.*, **38**, 465.

YATES, F. (1933*a*). The principles of orthogonality and confounding in replicated experiments. *J. Agr. Sci.*, **23**, 108.

YATES, F. (1933*b*). The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agr.*, **1**, 129.

YATES, F. (1933*c*). The formation of Latin squares for use in field experiments. *Emp. J. Exp. Agr.*, **1**, 235.

YATES, F. (1934*a*). The analysis of multiple classifications with unequal numbers in the different classes. *J. Am. Stat. Ass.*, **29**, 51.

YATES, F. (1934*b*). Contingency tables involving small numbers and the χ^2-test. *Supp. J.R.S.S.*, **1**, 217.

YATES, F. (1935*a*). Some examples of biassed sampling. *Ann. Eug. Lond.*, **6**, 202.

YATES, F. (1935*b*). Complex experiments. *Supp. J.R.S.S.*, **2**, 181.

YATES, F., and ZACOPANAY, I. (1935*c*). The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. *J. Agr. Sci.*, **25**, 545.

YATES, F. (1936*a*). Incomplete Latin squares. *J. Agr. Sci.*, **26**, 301.

YATES, F. (1936*b*). Incomplete randomised blocks. *Ann. Eug. Lond.*, **7**, 121.

YATES, F. (1936*c*). Applications of the sampling technique to crop estimation and forecasting. *Trans. Manchester Stat. Soc.*, **103**.

YATES, F. (1936*d*). A new method of arranging variety trials involving a large number of varieties. *J. Agr. Sci.*, **26**, 424.

YATES, F. (1937*a*). A further note on the arrangement of variety trials. Quasi-Latin squares. *Ann. Eug. Lond.*, **7**, 319.

YATES, F. (1937*b*). The design and analysis of factorial experiments. *Imp. Bur. Soil Sci. Tech. Comm.*, No. 35.

YATES, F. (1938*a*). The gain in efficiency resulting from the use of balanced designs. *Supp. J.R.S.S.*, **5**, 70.

YATES, F., and COCHRAN, W. G. (1938*b*). The analysis of groups of experiments. *J. Agr. Sci.*, **28**, 556.

YATES, F. (1938*c*). Orthogonal functions and tests of significance in the analysis of variance. *Supp. J.R.S.S.*, **5**, 177.

YATES, F. (1939*a*). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eug. Lond.*, **9**, 136.

YATES, F., and HALE, R. W. (1939*b*). The analysis of Latin squares when two or more rows, columns or treatments are missing. *Supp. J.R.S.S.*, **6**, 67.

YATES, F. (1939*c*). The adjustment of the weights of compound index numbers based on inaccurate data. *J.R.S.S.*, **102**, 285.

YATES, F. (1939*d*). Tests of significance of the differences between regression coefficients derived from two sets of correlated variates. *Proc. Roy. Soc. Edin.*, **59**, 184.

YATES, F. (1939*e*). The comparative advantages of systematic and randomised arrangements in the design of agricultural and biological experiments. *Biom.*, **30**, 440.

YATES, F. (1939*f*). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proc. Camb. Phil. Soc.*, **35**, 579.

YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eug. Lond.*, **10**, 317.

YOUNG, A. W., and PEARSON, K. (1916). On the probable error of a coefficient of contingency without approximation. *Biom.*, **11**, 215. (Correction, *Biom.*, **12**, 259.)

YOUNG, L. C. (1941). On randomness in ordered sequences. *Ann. Math. Stats.*, **12**, 293.

YULE, G. U. (1897*a*). On the significance of Bravais' formulæ for regression, etc., in the case of skew correlation. *Proc. Roy. Soc.*, **A**, **60**, 477.

YULE, G. U. (1897*b*). On the theory of correlation. *J.R.S.S.*, **60**, 812.

YULE, G. U. (1900). On the association of attributes etc. *Phil. Trans.*, **A**, **194**, 257.

YULE, G. U. (1906). On a property which holds good for all groupings of a normal distribution, etc. *Proc. Roy. Soc.*, **A**, **77**, 324.

YULE, G. U. (1907). On the theory of correlation for any number of variables treated by a new system of notation. *Proc. Roy. Soc.*, **A**, **79**, 182.

YULE, G. U. (1910). On the interpretation of correlations between indices or ratios. *J.R.S.S.*, **73**, 644.

YULE, G. U. (1912). On the methods of measuring the association between two attributes. *J.R.S.S.*, **75**, 579.

YULE, G. U. (1921). On the time-correlation problem. *J.R.S.S.*, **84**, 497.

YULE, G. U. (1922). On the application of the χ^2 method to association and contingency tables, with experimental illustrations. *J.R.S.S.*, **85**, 95.

YULE, G. U. (1926). Why do we sometimes get nonsense correlations between time-series, etc. ? *J.R.S.S.*, **89**, 1.

YULE, G. U. (1927*a*). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans.*, **A**, **226**, 267.

YULE, G. U. (1927*b*). On reading a scale. *J.R.S.S.*, **90**, 570.

YULE, G. U. (1936). On a parallelism between differential coefficients and regression coefficients. *J.R.S.S.*, **99**, 770.

YULE, G. U. (1938*a*). A test of Tippett's random sampling numbers. *J.R.S.S.*, **101**, 167.

YULE, G. U. (1938*b*). On some properties of normal distributions, univariate and bivariate, based on sums of squares of frequencies. *Biom.*, **30**, 1.

ZAYCOFF, R. (1936). Über die Zerlegung statistischer Zeitreihen in drei Komponenten. *Stat. Inst. Econ. Res. Univ. Sofia*, No. 4.

ZAYCOFF, R. (1937). Über die Ausschaltung der zufälligen Komponente nach der Variate-difference-Methode. *Stat. Inst. Econ. Res. Univ. Sofia*, No. 1.

ZIAUD-DIN, M. (1938). On differential operators developed by O'Toole. *Ann. Math. Stats.*, **9**, 63.

ZOCH, R. T. (1934). Invariants and covariants of certain frequency curves. *Ann. Math. Stats.*, **5**, 124.

ZOCH, R. T. (1935, 1937). On the postulate of the arithmetic mean. *Ann. Math. Stats.*, **6**, **171** ; and : Reply to Mr. Wertheimer's paper. *Ibid.*, **8**, 117.

ZRZAVY, F. J. (1933). Ausschaltung von Saisonschwankungen mittels Lag-correlation. *Monatsber. der Oest. Inst. für Konjunkturforschung*. Wien.

INDEX

(References are to pages. The abbreviations "*N.R.*" and "*Bibl.*" refer to the Notes and References and to the Bibliography respectively. Greek letters are indexed under their Roman equivalents, e.g. χ^2 under Chi-squared and ω under Omega.)